(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 December 2003 (18.12.2003)

PCT

(10) International Publication Number
WO 03/104410 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/17907

(22) International Filing Date: 5 June 2003 (05.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/387,119 7 June 2002 (07.06.2002) US
60/390,662 21 June 2002 (21.06.2002) US

(71) Applicant *(for all designated States except US)*: **INCYTE CORPORATION** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LEE, Soo, Y.** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). **CHAWLA, Narinder, K.** [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). **MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95037 (US). **HAFALIA, April, J.A.** [US/US]; 15 Midvale Drive, Daly City, CA 94015 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **BECHA, Shanya, D.** [US/US]; 1 Saint Francis #5508, San Francisco, CA 94107 (US). **BULLOCH, Sean, A.** [US/US]; 175 Calvert Drive #B-102, Cupertino, CA 95014 (US). **HAWKINS, Phillip, R.** [US/US]; 750 North Shorline Boulevard #115, Mountain View, CA 94043 (US). **JIN, Pei** [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **WILSON, Amy, D.** [US/US]; 1056 Continentals Way, #27, Belmont, CA 94002 (US). **CHANG, Hsin-Ru** [US/US]; 326 Treasure Island Drive, Belmont, CA 94002 (US). **KHARE, Reena** [IN/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **EMERLING, Brooke, M.** [US/US]; 833 W. Buena Avenue #2108, Chicago, IL 60613 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Corporation, 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/104410 A2

(54) **Title:** ENZYMES

(57) **Abstract:** Various embodiments of the invention provide human enzymes (ENZM) and polynucleotides which identify and encode ENZM. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of ENZM.

ENZYMES

TECHNICAL FIELD

The invention relates to novel nucleic acids, enzymes encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and enzymes.

BACKGROUND OF THE INVENTION

The cellular processes of biogenesis and biodegradation involve a number of key enzyme classes including oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, and others. Each class of enzyme comprises many substrate-specific enzymes having precise and well regulated functions. Enzymes facilitate metabolic processes such as glycolysis, the tricarboxylic cycle, and fatty acid metabolism; synthesis or degradation of amino acids, steroids, phospholipids, and alcohols; regulation of cell signaling, proliferation, inflamation, and apoptosis; and through catalyzing critical steps in DNA replication and repair and the process of translation.

<u>Oxidoreductases</u>

Many pathways of biogenesis and biodegradation require oxidoreductase (dehydrogenase or reductase) activity, coupled to reduction or oxidation of a cofactor. Potential cofactors include cytochromes, oxygen, disulfide, iron-sulfur proteins, flavin adenine dinucleotide (FAD), and the nicotinamide adenine dinucleotides NAD and NADP (Newsholme, E.A. and A.R. Leech (1983) <u>Biochemistry for the Medical Sciences</u>, John Wiley and Sons, Chichester, U. K. pp. 779-793). Reductase activity catalyzes transfer of electrons between substrate(s) and cofactor(s) with concurrent oxidation of the cofactor. Reverse dehydrogenase activity catalyzes the reduction of a cofactor and consequent oxidation of the substrate. Oxidoreductase enzymes are a broad superfamily that catalyze reactions in all cells of organisms, including metabolism of sugar, certain detoxification reactions, and synthesis or degradation of fatty acids, amino acids, glucocorticoids, estrogens, androgens, and prostaglandins. Different family members may be referred to as oxidoreductases, oxidases, reductases, or dehydrogenases, and they often have distinct cellular locations such as the cytosol, the plasma membrane, mitochondrial inner or outer membrane, and peroxisomes.

Short-chain alcohol dehydrogenases (SCADs) are a family of dehydrogenases that share only

15% to 30% sequence identity, with similarity predominantly in the coenzyme binding domain and the substrate binding domain. In addition to their role in detoxification of ethanol, SCADs are involved in synthesis and degradation of fatty acids, steroids, and some prostaglandins, and are therefore implicated in a variety of disorders such as lipid storage disease, myopathy, SCAD deficiency, and certain genetic disorders. For example, retinol dehydrogenase is a SCAD-family member (Simon, A. et al. (1995) J. Biol. Chem. 270:1107-1112) that converts retinol to retinal, the precursor of retinoic acid. Retinoic acid, a regulator of differentiation and apoptosis, has been shown to down-regulate genes involved in cell proliferation and inflammation (Chai, X. et al. (1995) J. Biol. Chem. 270:3900-3904). In addition, retinol dehydrogenase has been linked to hereditary eye diseases such as autosomal recessive childhood-onset severe retinal dystrophy (Simon, A. et al. (1996) Genomics 36:424-430).

Membrane-bound succinate dehydrogenases (succinate:quinone reductases, SQR) and fumarate reductases (quinol:fumarate reductases, QFR) couple the oxidation of succinate to fumarate with the reduction of quinone to quinol, and also catalyze the reverse reaction. QFR and SQR complexes are collectively known as succinate:quinone oxidoreductases (EC 1.3.5.1) and have similar compositions. The complexes consist of two hydrophilic and one or two hydrophobic, membrane-integrated subunits. The larger hydrophilic subunit A carries covalently bound flavin adenine dinucleotide; subunit B contains three iron-sulphur centers (Lancaster, C.R. and A. Kroger (2000) Biochim. Biophys. Acta 1459:422-431). The full-length cDNA sequence for the flavoprotein subunit of human heart succinate dehydrogenase (succinate: (acceptor) oxidoreductase; EC 1.3.99.1) is similar to the bovine succinate dehydrogenase in that it contains a cysteine triplet and in that the active site contains an additional cysteine that is not present in yeast or prokaryotic SQRs (Morris, A.A. et al. (1994) Biochim. Biophys. Acta 29:125-128).

Propagation of nerve impulses, modulation of cell proliferation and differentiation, induction of the immune response, and tissue homeostasis involve neurotransmitter metabolism (Weiss, B. (1991) Neurotoxicology 12:379-386; Collins, S.M. et al. (1992) Ann. N.Y. Acad. Sci. 664:415-424; Brown, J.K. and H. Imam (1991) J. Inherit. Metab. Dis. 14:436-458). Many pathways of neurotransmitter metabolism require oxidoreductase activity, coupled to reduction or oxidation of a cofactor, such as NAD^+/NADH (Newsholme and Leech, *supra*, pp. 779-793). Degradation of catecholamines (epinephrine or norepinephrine) requires alcohol dehydrogenase (in the brain) or aldehyde dehydrogenase (in peripheral tissue). NAD^+-dependent aldehyde dehydrogenase oxidizes 5-hydroxyindole-3-acetate (the product of 5-hydroxytryptamine (serotonin) metabolism) in the brain, blood platelets, liver and pulmonary endothelium (Newsholme and Leech, *supra*, p. 786). Other

neurotransmitter degradation pathways that utilize NAD^+/NADH-dependent oxidoreductase activity include those of L-DOPA (precursor of dopamine, a neuronal excitatory compound), glycine (an inhibitory neurotransmitter in the brain and spinal cord), histamine (liberated from mast cells during the inflammatory response), and taurine (an inhibitory neurotransmitter of the brain stem, spinal cord and retina) (Newsholme and Leech, *supra*, pp. 790, 792). Epigenetic or genetic defects in neurotransmitter metabolic pathways can result in diseases including Parkinson disease and inherited myoclonus (McCance, K.L. and S.E. Huether (1994) Pathophysiology, Mosby-Year Book, Inc., St. Louis, MO pp. 402-404; Gundlach, A.L. (1990) FASEB J. 4:2761-2766).

Tetrahydrofolate is a derivatized glutamate molecule that acts as a carrier, providing activated one-carbon units to a wide variety of biosynthetic reactions, including synthesis of purines, pyrimidines, and the amino acid methionine. Tetrahydrofolate is generated by the activity of a holoenzyme complex called tetrahydrofolate synthase, which includes three enzyme activities: tetrahydrofolate dehydrogenase, tetrahydrofolate cyclohydrolase, and tetrahydrofolate synthetase. Thus, tetrahydrofolate dehydrogenase plays an important role in generating building blocks for nucleic and amino acids, crucial to proliferating cells.

3-Hydroxyacyl-CoA dehydrogenase (3HACD) is involved in fatty acid metabolism. It catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA, with concomitant oxidation of NAD to NADH, in the mitochondria and peroxisomes of eukaryotic cells. In peroxisomes, 3HACD and enoyl-CoA hydratase form an enzyme complex called bifunctional enzyme, defects in which are associated with peroxisomal bifunctional enzyme deficiency. This interruption in fatty acid metabolism produces accumulation of very-long chain fatty acids, disrupting development of the brain, bone, and adrenal glands. Infants born with this deficiency typically die within 6 months (Watkins, P. et al. (1989) J. Clin. Invest. 83:771-777; Online Mendelian Inheritance in Man (OMIM), #261515). The neurodegeneration characteristic of Alzheimer's disease involves development of extracellular plaques in certain brain regions. A major protein component of these plaques is the peptide amyloid-β (Aβ), which is one of several cleavage products of amyloid precursor protein (APP). 3HACD has been shown to bind the Aβ peptide, and is overexpressed in neurons affected in Alzheimer's disease. In addition, an antibody against 3HACD can block the toxic effects of Aβ in a cell culture model of Alzheimer's disease (Yan, S. et al. (1997) Nature 389:689-695; OMIM, #602057).

Steroids such as estrogen, testosterone, and corticosterone are generated from a common precursor, cholesterol, and interconverted. Enzymes acting upon cholesterol include dehydrogenases. Steroid dehydrogenases, such as the hydroxysteroid dehydrogenases, are involved in hypertension, fertility, and cancer (Duax, W.L. and D. Ghosh (1997) Steroids 62:95-100). One such dehydrogenase

is 3-oxo-5-α-steroid dehydrogenase (OASD), a microsomal membrane protein highly expressed in prostate and other androgen-responsive tissues. OASD catalyzes the conversion of testosterone into dihydrotestosterone, which is the most potent androgen. Dihydrotestosterone is essential for the formation of the male phenotype during embryogenesis, as well as for proper androgen-mediated growth of tissues such as the prostate and male genitalia. A defect in OASD leads to defective formation of the external genitalia (Andersson, S. et al. (1991) Nature 354:159-161; Labrie, F. et al. (1992) Endocrinology 131:1571-1573; OMIM #264600).

17β-hydroxysteroid dehydrogenase (17βHSD6) plays an important role in the regulation of the male reproductive hormone, dihydrotestosterone (DHTT). 17βHSD6 acts to reduce levels of DHTT by oxidizing a precursor of DHTT, 3α-diol, to androsterone which is readily glucuronidated and removed. 17βHSD6 is active with both androgen and estrogen substrates in embryonic kidney 293 cells. Isozymes of 17βHSD catalyze oxidation and/or reduction reactions in various tissues with preferences for different steroid substrates (Biswas, M.G. and D.W. Russell (1997) J. Biol. Chem. 272:15959-15966). For example, 17βHSD1 preferentially reduces estradiol and is abundant in the ovary and placenta. 17βHSD2 catalyzes oxidation of androgens and is present in the endometrium and placenta. 17βHSD3 is exclusively a reductive enzyme in the testis (Geissler, W.M. et al. (1994) Nature Genet. 7:34-39). An excess of androgens such as DHTT can contribute to diseases such as benign prostatic hyperplasia and prostate cancer.

The oxidoreductase isocitrate dehydrogenase catalyzes the conversion of isocitrate to a-ketoglutarate, a substrate of the citric acid cycle. Isocitrate dehydrogenase can be either NAD or NADP dependent, and is found in the cytosol, mitochondria, and peroxisomes. Activity of isocitrate dehydrogenase is regulated developmentally, and by hormones, neurotransmitters, and growth factors.

Hydroxypyruvate reductase (HPR), a peroxisomal 2-hydroxyacid dehydrogenase in the glycolate pathway, catalyzes the conversion of hydroxypyruvate to glycerate with the oxidation of both NADH and NADPH. The reverse dehydrogenase reaction reduces NAD^+ and $NADP^+$. HPR recycles nucleotides and bases back into pathways leading to the synthesis of ATP and GTP, which are used to produce DNA and RNA and to control various aspects of signal transduction and energy metabolism. Purine nucleotide biosynthesis inhibitors are used as antiproliferative agents to treat cancer and viral diseases. HPR also regulates biochemical synthesis of serine and cellular serine levels available for protein synthesis.

The mitochondrial electron transport (or respiratory) chain is the series of oxidoreductase-type enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative

phosphorylation). ATP provides energy to drive energy-requiring reactions. The key respiratory chain complexes are NADH:ubiquinone oxidoreductase (complex I), succinate:ubiquinone oxidoreductase (complex II), cytochrome c_1-b oxidoreductase (complex III), cytochrome c oxidase (complex IV), and ATP synthase (complex V) (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, NY, pp. 677-678). All of these complexes are located on the inner matrix side of the mitochondrial membrane except complex II, which is on the cytosolic side where it transports electrons generated in the citric acid cycle to the respiratory chain. Electrons released in oxidation of succinate to fumarate in the citric acid cycle are transferred through electron carriers in complex II to membrane bound ubiquinone (Q). Transcriptional regulation of these nuclear-encoded genes controls the biogenesis of respiratory enzymes. Defects and altered expression of enzymes in the respiratory chain are associated with a variety of disease conditions.

Other dehydrogenase activities using NAD as a cofactor include 3-hydroxyisobutyrate dehydrogenase (3HBD), which catalyzes the NAD-dependent oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde within mitochondria. 3-hydroxyisobutyrate levels are elevated in ketoacidosis, methylmalonic acidemia, and other disorders (Rougraff, P.M. et al. (1989) J. Biol. Chem. 264:5899-5903). Another mitochondrial dehydrogenase important in amino acid metabolism is the enzyme isovaleryl-CoA-dehydrogenase (IVD). IVD is involved in leucine metabolism and catalyzes the oxidation of isovaleryl-CoA to 3-methylcrotonyl-CoA. Human IVD is a tetrameric flavoprotein synthesized in the cytosol with a mitochondrial import signal sequence. A mutation in the gene encoding IVD results in isovaleric acidemia (Vockley, J. et al. (1992) J. Biol. Chem. 267:2494-2501).

The family of glutathione peroxidases encompass tetrameric glutathione peroxidases (GPx1-3) and the monomeric phospholipid hydroperoxide glutathione peroxidase (PHGPx/GPx4). Although the overall homology between the tetrameric enzymes and GPx4 is less than 30%, a pronounced similarity has been detected in clusters involved in the active site and a common catalytic triad has been defined by structural and kinetic data (Epp, O. et al. (1983) Eur. J. Biochem. 133:51-69). GPx1 is ubiquitously expressed in cells, whereas GPx2 is present in the liver and colon, and GPx3 is present in plasma. GPx4 is found at low levels in all tissues but is expressed at high levels in the testis (Ursini, F. et al (1995) Meth. Enzymol. 252:38-53). GPx4 is the only monomeric glutathione peroxidase found in mammals and the only mammalian glutathione peroxidase to show high affinity for and reactivity with phospholipid hydroperoxides, and to be membrane associated. A tandem mechanism for the antioxidant activities of GPx4 and vitamin E has been suggested. GPx4 has alternative transcription and translation start sites which determine its subcellular localization (Esworthy, R.S. et al. (1994) Gene 144:317-318; and Maiorino, M. et al. (1990) Meth. Enzymol. 186:448-450).

The glutathione S-transferases (GST) are a ubiquitous family of enzymes with dual substrate specificities that perform important biochemical functions of xenobiotic biotransformation and detoxification, drug metabolism, and protection of tissues against peroxidative damage. They catalyze the conjugation of an electrophile with reduced glutathione (GSH) which results in either activation or deactivation/detoxification. The absolute requirement for binding reduced GSH to a variety of chemicals necessitates a diversity in GST structures in various organisms and cell types. GSTs are homodimeric or heterodimeric proteins localized in the cytosol. The major isozymes share common structural and catalytic properties and include four major classes, Alpha, Mu, Pi, and Theta. Each GST possesses a common binding site for GSH, and a variable hydrophobic binding site specific for its particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H.-C. et al. (1995) J. Biol. Chem. 270:99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg, G. et al. (1991) Biochem. J. 274:549-555).

GSTs normally deactivate and detoxify potentially mutagenic and carcinogenic chemicals. Some forms of rat and human GSTs are reliable preneoplastic markers of carcinogenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:8567-8580). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14:1371-1376). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents for which GST has affinity. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven, H.A. et al. (1994) Cancer Res. 54:6215-6220). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

The reduction of ribonucleotides to the corresponding deoxyribonucleotides, needed for DNA

synthesis during cell proliferation, is catalyzed by the enzyme ribonucleotide diphosphate reductase. Glutaredoxin is a glutathione (GSH)-dependent hydrogen donor for ribonucleotide diphosphate reductase and contains the active site consensus sequence -C-P-Y-C-. This sequence is conserved in glutaredoxins from such different organisms as *Escherichia coli*, vaccinia virus, yeast, plants, and mammalian cells. Glutaredoxin has inherent GSH-disulfide oxidoreductase (thioltransferase) activity in a coupled system with GSH, NADPH, and GSH-reductase, catalyzing the reduction of low molecular weight disulfides as well as proteins. Glutaredoxin has been proposed to exert a general thiol redox control of protein activity by acting both as an effective protein disulfide reductase, similar to thioredoxin, and as a specific GSH-mixed disulfide reductase (Padilla, C.A. et al. (1996) FEBS Lett. 378:69-73).

In addition to their important role in DNA synthesis and cell division, glutaredoxin and other thioproteins provide effective antioxidant defense against oxygen radicals and hydrogen peroxide (Schallreuter, K.U. and J.M. Wood (1991) Melanoma Res. 1:159-167). Glutaredoxin is the principal agent responsible for protein dethiolation *in vivo* and reduces dehydroascorbic acid in normal human neutrophils (Jung, C.H. and J.A. Thomas (1996) Arch. Biochem. Biophys. 335:61-72; Park, J.B. and M. Levine (1996) Biochem. J. 315:931-938).

The thioredoxin system serves as a hydrogen donor for ribonucleotide reductase and as a regulator of enzymes by redox control. It also modulates the activity of transcription factors such as NF-κB, AP-1, and steroid receptors. Several cytokines or secreted cytokine-like factors such as adult T-cell leukemia-derived factor, 3B6-interleukin-1, T-hybridoma-derived (MP-6) B cell stimulatory factor, and early pregnancy factor have been reported to be identical to thioredoxin (Holmgren, A. (1985) Annu. Rev. Biochem. 54:237-271; Abate, C. et al. (1990) Science 249:1157-1161; Tagaya, Y. et al. (1989) EMBO J. 8:757-764; Wakasugi, H. (1987) Proc. Natl. Acad. Sci. USA 84:804-808; Rosen, A. et al. (1995) Int. Immunol. 7:625-633). Thus thioredoxin secreted by stimulated lymphocytes (Yodoi, J. and T. Tursz (1991) Adv. Cancer Res. 57:381-411; Tagaya, N. et al. (1990) Proc. Natl. Acad. Sci. USA 87:8282-8286) has extracellular activities including a role as a regulator of cell growth and a mediator in the immune system (Miranda-Vizuete, A. et al. (1996) J. Biol. Chem. 271:19099-19103; Yamauchi, A. et al. (1992) Mol. Immunol. 29:263-270). Thioredoxin and thioredoxin reductase protect against cytotoxicity mediated by reactive oxygen species in disorders such as Alzheimer's disease (Lovell, M.A. (2000) Free Radic. Biol. Med. 28:418-427).

The selenoprotein thioredoxin reductase is secreted by both normal and neoplastic cells and has been implicated as both a growth factor and as a polypeptide involved in apoptosis (Soderberg, A. et al. (2000) Cancer Res. 60:2281-2289). An extracellular plasmin reductase secreted by hamster

ovary cells (HT-1080) has been shown to participate in the generation of angiostatin from plasmin. In this case, the reduction of the plasmin disulfide bonds triggers the proteolytic cleavage of plasmin which yields the angiogenesis inhibitor, angiostatin (Stathakis, P. et al. (1997) J. Biol. Chem. 272:20641-20645). Low levels of reduced sulfhydryl groups in plasma has been associated with rheumatoid arthritis. The failure of these sulfhydryl groups to scavenge active oxygen species (e.g., hydrogen peroxide produced by activated neutrophils) results in oxidative damage to surrounding tissues and the resulting inflammation (Hall, N.D. et al. (1994) Rheumatol. Int. 4:35-38).

Another example of the importance of redox reactions in cell metabolism is the degradation of saturated and unsaturated fatty acids by mitochondrial and peroxisomal beta-oxidation enzymes which sequentially remove two-carbon units from Coenzyme A (CoA)-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids.

The pathways of mitchondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) Biochem. J. 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) Biochimie 75:147-158).

The auxiliary beta-oxidation enzyme 2,4-dienoyl-CoA reductase catalyzes the following reaction:

trans-2, cis/trans-4-dienoyl-CoA + NADPH + H^+ ---> trans-3-enoyl-CoA + $NADP^+$

This reaction removes even-numbered double bonds from unsaturated fatty acids prior to their entry into the main beta-oxidation pathway (Koivuranta, K.T. et al. (1994) Biochem. J. 304:787-792). The enzyme may also remove odd-numbered double bonds from unsaturated fatty acids (Smeland, T.E. et al. (1992) Proc. Natl. Acad. Sci. USA 89:6673-6677).

Rat 2,4-dienoyl-CoA reductase is located in both mitochondria and peroxisomes (Dommes, V. et al. (1981) J. Biol. Chem. 256:8259-8262). Two immunologically different forms of rat mitochondrial enzyme exist with molecular masses of 60 kDa and 120 kDa (Hakkola, E.H. and J.K.

Hiltunen (1993) Eur. J. Biochem. 215:199-204). The 120 kDa mitochondrial rat enzyme is synthesized as a 335 amino acid precursor with a 29 amino acid N-terminal leader peptide which is cleaved to form the mature enzyme (Hirose, A. et al. (1990) Biochim. Biophys. Acta 1049:346-349). A human mitochondrial enzyme 83% similar to rat enzyme is synthesized as a 335 amino acid residue precursor with a 19 amino acid N-terminal leader peptide (Koivuranta et al., *supra*). These cloned human and rat mitochondrial enzymes function as homotetramers (Koivuranta et al., *supra*). A *Saccharomyces cerevisiae* peroxisomal 2,4-dienoyl-CoA reductase is 295 amino acids long, contains a C-terminal peroxisomal targeting signal, and functions as a homodimer (Coe, J.G.S. et al. (1994) Mol. Gen. Genet. 244:661-672; and Gurvitz, A. et al. (1997) J. Biol. Chem. 272:22140-22147). All 2,4-dienoyl-CoA reductases have a fairly well conserved NADPH binding site motif (Koivuranta et al., *supra*).

The main pathway beta-oxidation enzyme enoyl-CoA hydratase catalyzes the reaction:

2-trans-enoyl-CoA + H_2O <---> 3-hydroxyacyl-CoA

This reaction hydrates the double bond between C-2 and C-3 of 2-trans-enoyl-CoA, which is generated from saturated and unsaturated fatty acids (Engel, C.K. et al. (1996) EMBO J. 15:5135-5145). This step is downstream from the step catalyzed by 2,4-dienoyl-reductase. Different enoyl-CoA hydratases act on short-, medium-, and long-chain fatty acids (Eaton et al., *supra*). Mitochondrial and peroxisomal enoyl-CoA hydratases occur as both mono-functional enzymes and as part of multi-functional enzyme complexes. Human liver mitochondrial short-chain enoyl-CoA hydratase is synthesized as a 290 amino acid precursor with a 29 amino acid N-terminal leader peptide (Kanazawa, M. et al. (1993) Enzyme Protein 47:9-13; and Janssen, U. et al. (1997) Genomics 40:470-475). Rat short-chain enoyl-CoA hydratase is 87% identical to the human sequence in the mature region of the protein and functions as a homohexamer (Kanazawa et al., *supra*; and Engel et al., *supra*). A mitochondrial trifunctional protein exists that has long-chain enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and long-chain 3-oxothiolase activities (Eaton et al., *supra*). In human peroxisomes, enoyl-CoA hydratase activity is found in both a 327 amino acid residue mono-functional enzyme and as part of a multi-functional enzyme, also known as bifunctional enzyme, which possesses enoyl-CoA hydratase, enoyl-CoA isomerase, and 3-hydroxyacyl-CoA hydrogenase activities (FitzPatrick, D.R. et al. (1995) Genomics 27:457-466; and Hoefler, G. et al. (1994) Genomics 19:60-67). A 339 amino acid residue human protein with short-chain enoyl-CoA hydratase activity also acts as an AU-specific RNA binding protein (Nakagawa, J. et al. (1995) Proc. Natl. Acad. Sci. USA 92:2051-2055). All enoyl-CoA hydratases share homology near two active site glutamic acid

residues, with 17 amino acid residues that are highly conserved (Wu, W.-J. et al. (1997) Biochemistry 36:2211-2220).

Inherited deficiencies in mitochondrial and peroxisomal beta-oxidation enzymes are associated with severe diseases, some of which manifest soon after birth and lead to death within a few years. Mitochondrial beta-oxidation associated deficiencies include, e.g., carnitine palmitoyl transferase and carnitine deficiency, very-long-chain acyl-CoA dehydrogenase deficiency, medium-chain acyl-CoA dehydrogenase deficiency, short-chain acyl-CoA dehydrogenase deficiency, electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, trifunctional protein deficiency, and short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (Eaton et al., *supra*). Mitochondrial trifunctional protein (including enoyl-CoA hydratase) deficient patients have reduced long-chain enoyl-CoA hydratase activities and suffer from non-ketotic hypoglycemia, sudden infant death syndrome, cardiomyopathy, hepatic dysfunction, and muscle weakness, and may die at an early age (Eaton et al., *supra*).

Defects in mitochondrial beta-oxidation are associated with Reye's syndrome, a disease characterized by hepatic dysfunction and encephalopathy that sometimes follows viral infection in children. Reye's syndrome patients may have elevated serum levels of free fatty acids (Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA, p.866). Patients with mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency and medium-chain 3-hydroxyacyl-CoA dehydrogenase deficiency also exhibit Reye-like illnesses (Eaton et al., *supra*; and Egidio, R.J. et al. (1989) Am. Fam. Physician 39:221-226).

Inherited conditions associated with peroxisomal beta-oxidation include Zellweger syndrome, neonatal adrenoleukodystrophy, infantile Refsum's disease, acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, and bifunctional protein deficiency (Suzuki, Y. et al. (1994) Am. J. Hum. Genet. 54:36-43; Hoefler et al., *supra*). Patients with peroxisomal bifunctional enzyme deficiency, including that of enoyl-CoA hydratase, suffer from hypotonia, seizures, psychomotor defects, and defective neuronal migration; accumulate very-long-chain fatty acids; and typically die within a few years of birth (Watkins, P.A. et al. (1989) J. Clin. Invest. 83:771-777).

Peroxisomal beta-oxidation is impaired in cancerous tissue. Although neoplastic human breast epithelial cells have the same number of peroxisomes as do normal cells, fatty acyl-CoA oxidase activity is lower than in control tissue (el Bouhtoury, F. et al. (1992) J. Pathol. 166:27-35). Human colon carcinomas have fewer peroxisomes than normal colon tissue and have lower fatty-acyl-CoA oxidase and bifunctional enzyme (including enoyl-CoA hydratase) activities than normal tissue (Cable, S. et al. (1992) Virchows Arch. B Cell Pathol. Incl. Mol. Pathol. 62:221-226).

6-phosphogluconate dehydrogenase (6-PGDH) catalyses the $NADP^+$-dependent oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate with the production of NADPH. The absence or inhibition of 6-PGDH results in the accumulation of 6-phosphogluconate to toxic levels in eukaryotic cells. 6-PGDH is the third enzyme of the pentose phosphate pathway (PPP) and is ubiquitous in nature. In some heterofermentatative species, NAD+ is used as a cofactor with the subsequent production of NADH.

The reaction proceeds through a 3-keto intermediate which is decarboxylated to give the enol of ribulose 5-phosphate, then converted to the keto product following tautomerization of the enol (Berdis A.J. and P.F. Cook (1993) Biochemistry 32:2041-2046). 6-PGDH activity is regulated by the inhibitory effect of NADPH, and the activating effect of 6-phosphogluconate (Rippa, M. et al. (1998) Biochim. Biophys. Acta 1429:83-92). Deficiencies in 6-PGDH activity have been linked to chronic hemolytic anemia.

The targeting of specific forms of 6-PGDH (e.g., enzymes found in trypanosomes) has been suggested as a means for controlling parasitic infections (Tetaud, E. et al. (1999) Biochem. J. 338:55-60). For example, the *Trypanosoma brucei* enzyme is markedly more sensitive to inhibition by the substrate analogue 6-phospho-2-deoxygluconate and the coenzyme analogue adenosine 2',5'-bisphosphate, compared to the mammalian enzyme (Hanau, S. et al. (1996) Eur. J. Biochem. 240:592-599).

Ribonucleotide diphosphate reductase catalyzes the reduction of ribonucleotide diphosphates (i.e., ADP, GDP, CDP, and UDP) to their corresponding deoxyribonucleotide diphosphates (i.e., dADP, dGDP, dCDP, and dUDP) which are used for the synthesis of DNA. Ribonucleotide diphosphate reductase thereby performs a crucial role in the *de novo* synthesis of deoxynucleotide precursors. Deoxynucleotides are also produced from deoxynucleosides by nucleoside kinases via the salvage pathway.

Mammalian ribonucleotide diphosphate reductase comprises two components, an effector-binding component (E) and a non-heme iron component (F). Component E binds the nucleoside triphosphate effectors while component F contains the iron radical necessary for catalysis. Molecular weight determinations of the E and F components, as well as the holoenzyme, vary according to the methods used in purification of the proteins and the particular laboratory. Component E is approximately 90-100 kDa, component F is approximately 100-120 kDa, and the holoenzyme is 200-250 kDa.

Ribonucleotide diphosphate reductase activity is adversely effected by iron chelators, such as thiosemicarbazones, as well as EDTA. Deoxyribonucleotide diphosphates also appear to be negative

allosteric effectors of ribonucleotide diphosphate reductase. Nucleotide triphosphates (both ribo- and deoxyribo-) appear to stimulate the activity of the enzyme. 3-methyl-4-nitrophenol, a metabolite of widely used organophosphate pesticides, is a potent inhibitor of ribonucleotide diphosphate reductase in mammalian cells. Some evidence suggests that ribonucleotide diphosphate reductase activity in DNA virus (e.g., herpes virus) -infected cells and in cancer cells is less sensitive to regulation by allosteric regulators and a correlation exists between high ribonucleotide diphosphate reductase activity levels and high rates of cell proliferation (e.g., in hepatomas). This observation suggests that virus-encoded ribonucleotide diphosphate reductases, and those present in cancer cells, are capable of maintaining an increased supply deoxyribonucleotide pool for the production of virus genomes or for the increased DNA synthesis which characterizes cancers cells. Ribonucleotide diphosphate reductase is thus a target for therapeutic intervention (Nutter, L.M. and Y.-C. Cheng (1984) Pharmac. Ther. 26:191-207; and Wright, J.A. (1983) Pharmac. Ther. 22:81-102).

Dihydrodiol dehydrogenases (DD) are monomeric, $NAD(P)^+$-dependent, 34-37 kDa enzymes responsible for the detoxification of *trans*-dihydrodiol and *anti*-diol epoxide metabolites of polycyclic aromatic hydrocarbons (PAH) such as benzo[*a*]yrene, benz[*a*]anthracene, 7-methyl-benz[*a*]anthracene, 7,12-dimethyl-benz[*a*]anthracene, chrysene, and 5-methyl-chrysene. In mammalian cells, an environmental PAH toxin such as benzo[*a*]yrene is initially epoxidated by a microsomal cytochrome P450 to yield 7*R*,8*R*-arene-oxide and subsequently (-)-7*R*,8*R*-dihydrodiol ((-)-*trans*-7,8-dihydroxy-7,8-dihydrobenzo[*a*]pyrene or (-)-*trans*-B[*a*]P-diol) This latter compound is further transformed to the *anti*-diol epoxide of benzo[*a*]pyrene (i.e., (±)-anti-7β,8α-dihydroxy-9α,10α-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene), by the same enzyme or a different enzyme, depending on the species. This resulting *anti*-diol epoxide of benzo[*a*]yrene, or the corresponding derivative from another PAH compound, is highly mutagenic.

DD efficiently oxidizes the precursor of the *anti*-diol epoxide (i.e., *trans*-dihydrodiol) to transient catechols which auto-oxidize to quinones, also producing hydrogen peroxide and semiquinone radicals. This reaction prevents the formation of the highly carcinogenic *anti*-diol. *Anti*-diols are not themselves substrates for DD yet the addition of DD to a sample comprising an *anti*-diol compound results in a significant decrease in the induced mutation rate observed in the Ames test. In this instance, DD is able to bind to and sequester the *anti*-diol, even though it is not oxidized. Whether through oxidation or sequestration, DD plays an important role in the detoxification of metabolites of xenobiotic polycyclic compounds (Penning, T.M. (1993) Chemico-Biological Interactions 89:1-34).

15-oxoprostaglandin 13-reductase (PGR) and 15-hydroxyprostaglandin dehydrogenase (15-PGDH) are enzymes present in the lung that are responsible for degrading circulating prostaglandins.

Oxidative catabolism via passage through the pulmonary system is a common means of reducing the concentration of circulating prostaglandins. 15-PGDH oxidizes the 15-hydroxyl group of a variety of prostaglandins to produce the corresponding 15-oxo compounds. The 15-oxo derivatives usually have reduced biological activity compared to the 15-hydroxyl molecule. PGR further reduces the 13,14 double bond of the 15-oxo compound which typically leads to a further decrease in biological activity. PGR is a monomer with a molecular weight of approximately 36 kDa. The enzyme requires NADH or NADPH as a cofactor with a preference for NADH. The 15-oxo derivatives of prostaglandins PGE_1, PGE_2, and $PGE_{2\alpha}$ are all substrates for PGR; however, the non-derivatized prostaglandins (i.e., PGE_1, PGE_2, and $PGE_{2\alpha}$) are not substrates (Ensor, C.M. et al. (1998) Biochem. J. 330:103-108).

15-PGDH and PGR also catalyze the metabolism of lipoxin A_4 (LXA_4). Lipoxins (LX) are autacoids, lipids produced at the sites of localized inflammation, which down-regulate polymorphonuclear leukocyte (PMN) function and promote resolution of localized trauma. Lipoxin production is stimulated by the administration of aspirin in that cells displaying cyclooxygenase II (COX II) that has been acetylated by aspirin and cells that possess 5-lipoxygenase (5-LO) interact and produce lipoxin. 15-PGDH generates 15-oxo-LXA_4 with PGR further converting the 15-oxo compound to 13,14-dihydro-15-oxo-LXA_4 (Clish, C.B. et al. (2000) J. Biol. Chem. 275:25372-25380). This finding suggests a broad substrate specificity of the prostaglandin dehydrogenases and has implications for these enzymes in drug metabolism and as targets for therapeutic intervention to regulate inflammation.

The GMC (glucose–methanol–choline) oxidoreductase family of enzymes was defined based on sequence alignments of *Drosophila melanogaster* glucose dehydrogenase, *Escherichia coli* choline dehydrogenase, *Aspergillus niger* glucose oxidase, and *Hansenula polymorpha* methanol oxidase. Despite their different sources and substrate specificities, these four flavoproteins are homologous, being characterized by the presence of several distinctive sequence and structural features. Each molecule contains a canonical ADP-binding, beta-alpha-beta mononucleotide-binding motif close to the amino terminus. This fold comprises a four-stranded parallel beta-sheet sandwiched between a three-stranded antiparallel beta-sheet and alpha-helices. Nucleotides bind in similar positions relative to this chain fold (Cavener, D.R. (1992) J. Mol. Biol. 223:811-814; Wierenga, R.K. et al. (1986) J. Mol. Biol. 187:101-107). Members of the GMC oxidoreductase family also share a consensus sequence near the central region of the polypeptide. Additional members of the GMC oxidoreductase family include cholesterol oxidases from *Brevibacterium sterolicum* and *Streptomyces*; and an alcohol dehydrogenase from *Pseudomonas oleovorans* (Cavener, *supra*; Henikoff, S. and J.G. Henikoff (1994) Genomics 19:97-107; van Beilen, J.B. et al. (1992) Mol.

Microbiol. 6:3121-3136).

IMP dehydrogenase and GMP reductase are two oxidoreductases which share many regions of sequence similarity. IMP dehydrogenase (EC 1.1.1.205) catalyes the NAD-dependent reduction of IMP (inosine monophosphate) into XMP (xanthine monophosphate) as part of *de novo* GTP biosynthesis (Collart, F.R. and E. Huberman (1988) J. Biol. Chem. 263:15769-15772). GMP reductase catalyzes the NADPH-dependent reductive deamination of GMP into IMP, helping to maintain the intracellular balance of adenine and guanine nucleotides (Andrews, S.C. and J.R. Guest (1988) Biochem. J. 255:35-43).

Pyridine nucleotide-disulphide oxidoreductases are FAD flavoproteins involved in the transfer of reducing equivalents from FAD to a substrate. These flavoproteins contain a pair of redox-active cysteines contained within a consensus sequence which is characteristic of this protein family (Kurlyan, J. et al. (1991) Nature 352:172-174). Members of this family of oxidoreductases include glutathione reductase (EC 1.6.4.2); thioredoxin reductase of higher eukaryotes (EC 1.6.4.5); trypanothione reductase (EC 1.6.4.8); lipoamide dehydrogenase (EC 1.8.1.4), the E3 component of alpha-ketoacid dehydrogenase complexes; and mercuric reductase (EC 1.16.1.1).

Transferases

Transferases are enzymes that catalyze the transfer of molecular groups. The reaction may involve an oxidation, reduction, or cleavage of covalent bonds, and is often specific to a substrate or to particular sites on a type of substrate. Transferases participate in reactions essential to such functions as synthesis and degradation of cell components, and regulation of cell functions including cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Transferases are involved in key steps in disease processes involving these functions. Transferases are frequently classified according to the type of group transferred. For example, methyl transferases transfer one-carbon methyl groups, amino transferases transfer nitrogenous amino groups, and similarly denominated enzymes transfer aldehyde or ketone, acyl, glycosyl, alkyl or aryl, isoprenyl, saccharyl, phosphorous-containing, sulfur-containing, or selenium-containing groups, as well as small enzymatic groups such as Coenzyme A.

Acyl transferases include peroxisomal carnitine octanoyl transferase, which is involved in the fatty acid beta-oxidation pathway, and mitochondrial carnitine palmitoyl transferases, involved in fatty acid metabolism and transport. Choline O-acetyl transferase catalyzes the biosynthesis of the neurotransmitter acetylcholine. N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are

then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C.N. et al. (1994) J. Biol. Chem. 269:19375-19379; Johnson, M.R. et al. (1991) J. Biol. Chem. 266:10227-10233). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-1445).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from *Saccharomyces cerevisiae*. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Curr. Opin. Cell Biol. 12:326-333 and Berger, S.L (1999) Curr. Opin. Cell Biol. 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, http://predict.sanger.ac.uk/irbm-course97/Docs/ms/) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group to aromatic amines and hydrazine containing compounds. In humans, there are two highly similar N-acetyltransferase enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping

substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminoglutethimide, and sulfamethazine). A recently isolated human gene, tubedown-1, is homologous to the yeast NAT-1 N-acetyltransferases and encodes a protein associated with acetyltransferase activity. The expression patterns of tubedown-1 suggest that it may be involved in regulating vascular and hematopoietic development (Gendron, R.L. et al. (2000) Dev. Dyn. 218:300-315).

Amino transferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Amino transferases play key roles in protein synthesis and degradation, and they contribute to other processes as well. For example, GABA aminotransferase (GABA-T) catalyzes the degradation of GABA, the major inhibitory amino acid neurotransmitter. The activity of GABA-T is correlated to neuropsychiatric disorders such as alcoholism, epilepsy, and Alzheimer's disease (Sherif, F.M. and S.S. Ahmed (1995) Clin. Biochem. 28:145-154). Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937). Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyzes the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleiotropic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Glycosyl transferases include the mammalian UDP-glucouronosyl transferases, a family of membrane-bound microsomal enzymes catalyzing the transfer of glucouronic acid to lipophilic substrates in reactions that play important roles in detoxification and excretion of drugs, carcinogens, and other foreign substances. Another mammalian glycosyl transferase, mammalian UDP-galactose-ceramide galactosyl transferase, catalyzes the transfer of galactose to ceramide in the synthesis of galactocerebrosides in myelin membranes of the nervous system. The UDP-glycosyl transferases share a conserved signature domain of about 50 amino acid residues (PROSITE: PDOC00359, http://expasy.hcuge.ch/sprot/prosite.html).

Methyl transferases are involved in a variety of pharmacologically important processes. Nicotinamide N-methyl transferase catalyzes the N-methylation of nicotinamides and other pyridines, an important step in the cellular handling of drugs and other foreign compounds. Phenylethanolamine

N-methyl transferase catalyzes the conversion of noradrenalin to adrenalin. 6-O-methylguanine-DNA methyl transferase reverses DNA methylation, an important step in carcinogenesis. Uroporphyrin-III C-methyl transferase, which catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to uroporphyrinogen III, is the first specific enzyme in the biosynthesis of cobalamin, a dietary enzyme whose uptake is deficient in pernicious anemia. Protein-arginine methyl transferases catalyze the posttranslational methylation of arginine residues in proteins, resulting in the mono- and dimethylation of arginine on the guanidino group. Substrates include histones, myelin basic protein, and heterogeneous nuclear ribonucleoproteins involved in mRNA processing, splicing, and transport. Protein-arginine methyl transferase interacts with proteins upregulated by mitogens, with proteins involved in chronic lymphocytic leukemia, and with interferon, suggesting an important role for methylation in cytokine receptor signaling (Lin, W.-J. et al. (1996) J. Biol. Chem. 271:15034-15044; Abramovich, C. et al. (1997) EMBO J. 16:260-266; and Scott, H. S. et al. (1998) Genomics 48:330-340).

Phospho transferases catalyze the transfer of high-energy phosphate groups and are important in energy-requiring and -releasing reactions. The metabolic enzyme creatine kinase catalyzes the reversible phosphate transfer between creatine/creatine phosphate and ATP/ADP. Glycocyamine kinase catalyzes phosphate transfer from ATP to guanidoacetate, and arginine kinase catalyzes phosphate transfer from ATP to arginine. A cysteine-containing active site is conserved in this family (PROSITE: PDOC00103).

Prenyl transferases are heterodimers, consisting of an alpha and a beta subunit, that catalyze the transfer of an isoprenyl group. The Ras farnesyltransferase (FTase) enzyme transfers a farnesyl moiety from cytosolic farnesylpyrophosphate to a cysteine residue at the carboxyl terminus of the Ras oncogene protein. This modification is required to anchor Ras to the cell membrane so that it can perform its role in signal transduction. FTase inhibitors block Ras function and demonstrate antitumor activity_(Buolamwini, J.K. (1999) Curr. Opin. Chem. Biol. 3:500-509). Ftase, which shares structural similarity with geranylgeranyl transferase, or Rab GG transferase, prenylates Rab proteins, allowing them to perform their roles in regulating vesicle transport (Seabra, M.C. (1996) J. Biol. Chem. 271:14398-14404).

Saccharyl transferases are glycating enzymes involved in a variety of metabolic processes. Oligosaccharyl transferase-48, for example, is a receptor for advanced glycation endproducts, which accumulate in vascular complications of diabetes, macrovascular disease, renal insufficiency, and Alzheimer's disease (Thornalley, P. J. (1998) Cell Mol. Biol. (Noisy-Le-Grand) 44:1013-1023).

Coenzyme A (CoA) transferase catalyzes the transfer of CoA between two carboxylic acids.

Succinyl CoA:3-oxoacid CoA transferase, for example, transfers CoA from succinyl-CoA to a recipient such as acetoacetate. Acetoacetate is essential to the metabolism of ketone bodies, which accumulate in tissues affected by metabolic disorders such as diabetes (PROSITE: PDOC00980).

Transglutaminase transferases (Tgases) are Ca^{2+} dependent enzymes capable of forming isopeptide bonds by catalyzing the transfer of the γ-carboxy group from protein-bound glutamine to the ε-amino group of protein-bound lysine residues or other primary amines. Tgases are the enzymes responsible for the cross-linking of cornified envelope (CE), the highly insoluble protein structure on the surface of corneocytes, into a chemically and mechanically resistant protein polymer. Seven known human Tgases have been identified. Individual transglutaminase gene products are specialized in the cross-linking of specific proteins or tissue structures, such as factor XIIIa which stabilizes the fibrin clot in hemostasis, prostrate transglutaminase which functions in semen coagulation, and tissue transglutaminase which is involved in GTP-binding in receptor signaling. Four (Tgases 1, 2, 3, and X) are expressed in terminally differentiating epithelia such as the epidermis. Tgases are critical for the proper cross-linking of the CE as seen in the pathology of patients suffering from one form of the skin diseases referred to as congenital ichthyosis which has been linked to mutations in the keratinocyte transglutaminase (TG_K) gene (Nemes, Z. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:8402-8407, Aeschlimann, D. et al. (1998) J. Biol. Chem. 273:3452-3460.)

Hydrolases

Hydrolases are a class of enzymes that catalyze the cleavage of various covalent bonds in a substrate by the introduction of a molecule of water. The reaction involves a nucleophilic attack by the water molecule's oxygen atom on a target bond in the substrate. The water molecule is split across the target bond, breaking the bond and generating two product molecules. Hydrolases participate in reactions essential to such functions as synthesis and degradation of cell components, and for regulation of cell functions including cell signaling, cell proliferation, inflamation, apoptosis, secretion and excretion. Hydrolases are involved in key steps in disease processes involving these functions. Hydrolytic enzymes, or hydrolases, may be grouped by substrate specificity into classes including phosphatases, peptidases, lysophospholipases, phosphodiesterases, glycosidases, glyoxalases, aminohydrolases, carboxylesterases, sulfatases, phosphohydrolases, nucleotidases, lysozymes, and many others.

Phosphatases hydrolytically remove phosphate groups from proteins, an energy-providing step that regulates many cellular processes, including intracellular signaling pathways that in turn control cell growth and differentiation, cell-cell contact, the cell cycle, and oncogenesis.

Peptidases, also called proteases, cleave peptide bonds that form the backbone of peptide or protein chains. Proteolytic processing is essential to cell growth, differentiation, remodeling, and homeostasis as well as inflammation and the immune response. Since typical protein half-lives range from hours to a few days, peptidases are continually cleaving precursor proteins to their active form, removing signal sequences from targeted proteins, and degrading aged or defective proteins. Peptidases function in bacterial, parasitic, and viral invasion and replication within a host. Examples of peptidases include trypsin and chymotrypsin (components of the complement cascade and the blood-clotting cascade) lysosomal cathepsins, calpains, pepsin, renin, and chymosin (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp. 1-5).

Lysophospholipases (LPLs) regulate intracellular lipids by catalyzing the hydrolysis of ester bonds to remove an acyl group, a key step in lipid degradation. Small LPL isoforms, approximately 15-30 kD, function as hydrolases; larger isoforms function both as hydrolases and transacylases. A particular substrate for LPLs, lysophosphatidylcholine, causes lysis of cell membranes. LPL activity is regulated by signaling molecules important in numerous pathways, including the inflammatory response.

The phosphodiesterases catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endo- and exo-nucleases, which are essential to cell growth and replication as well as protein synthesis. Endonuclease V (deoxyinosine 3′-endonuclease) is an example of a type II site-specific deoxyribonuclease, a putative DNA repair enzyme that cleaves DNAs containing hypoxanthine, uracil, or mismatched bases. *Escherichia coli* endonuclease V has been shown to cleave DNA containing deoxyxanthosine at the second phosphodiester bond 3′ to deoxyxanthosine, generating a 3′-hydroxyl and a 5′-phosphoryl group at the nick site (He, B. et al. (2000) Mutat. Res. 459:109-114). It has been suggested that *Escherichia coli* endonuclease V plays a role in the removal of deaminated guanine, i.e., xanthine, from DNA, thus helping to protect the cell against the mutagenic effects of nitrosative deamination (Schouten, K.A. and B. Weiss (1999) Mutat. Res. 435:245-254). In eukaryotes, the process of tRNA splicing requires the removal of small tRNA introns that interrupt the anticodon loop 1 base 3′ to the anticodon. This process requires the stepwise action of an endonuclease, a ligase, and a phosphotransferase (Hong, L. et al. (1998) Science 280:279-284). Ribonuclease P (RNase P) is a ubiquitous RNA processing endonuclease that is required for generating the mature tRNA 5′-end during the tRNA splicing process. This is accomplished through the catalysis of the cleavage of P-3′O bonds to produce 5′-phosphate and 3′-hydroxyl end groups at

a specific site on pre-tRNA. Catalysis by RNase P is absolutely dependent on divalent cations such as Mg^{2+} or Mn^{2+} (Kurz, J.C. et al. (2000) Curr. Opin. Chem. Biol. 4:553-558). Substrate recognition mechanisms of RNase P are well conserved among eukaryotes and bacteria (Fabbri, S. et al. (1998) Science 280:284-286). In *Saccharomyces cerevisiae*, POP1 ('processing of precursor RNAs') encodes a protein component of both RNase P and RNase MRP, another RNA processing protein. Mutations in yeast POP1 are lethal (Lygerou, Z. et al. (1994) Genes Dev. 8:1423-1433). Another phosphodiesterase, acid sphingomyelinase, hydrolyzes the membrane phospholipid sphingomyelin to ceramide and phosphorylcholine. Phosphorylcholine functions in synthesis of phosphatidylcholine, which is involved in intracellular signaling pathways. Ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase phosphodiesterase leads to Niemann-Pick disease.

Glycosidases catalyze the cleavage of hemiacetyl bonds of glycosides, which are compounds that contain one or more sugar. Mammalian lactase-phlorizin hydrolase, for example, is an intestinal enzyme that splits lactose. Mammalian beta-galactosidase removes the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans, and deficiency of this enzyme is associated with a gangliosidosis known as Morquio disease type B (PROSITE PCDOC00910). Vertebrate lysosomal alpha-glucosidase, which hydrolyzes glycogen, maltose, and isomaltose, and vertebrate intestinal sucrase-isomaltase, which hydrolyzes sucrose, maltose, and isomaltose, are widely distributed members of this family with highly conserved sequences at their active sites.

The glyoxylase system is involved in gluconeogenesis, the production of glucose from storage compounds in the body. It consists of glyoxylase I, which catalyzes the formation of S-D-lactoylglutathione from methyglyoxal, a side product of triose-phosphate energy metabolism, and glyoxylase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Glyoxylases are involved in hyperglycemia, non-insulin-dependent diabetes mellitus, the detoxification of bacterial toxins, and in the control of cell proliferation and microtubule assembly.

NG,NG-dimethylarginine dimethylaminohydrolase (DDAH) is an enzyme that hydrolyzes the endogenous nitric oxide synthase (NOS) inhibitors, NG-monomethyl-arginine and NG,NG-dimethyl-L-arginine, to L-citrulline. Inhibiting DDAH can cause increased intracellular concentration of NOS inhibitors to levels sufficient to inhibit NOS. Therefore, DDAH inhibition may provide a method of NOS inhibition, and changes in the activity of DDAH could play a role in pathophysiological alterations in nitric oxide generation (MacAllister, R.J. et al. (1996) Br. J. Pharmacol. 119:1533-1540). DDAH was found in neurons displaying cytoskeletal abnormalities and oxidative stress in Alzheimer's disease. In age-matched control cases, DDAH was not found in neurons. This suggests that oxidative stress-

and nitric oxide-mediated events play a role in the pathogenesis of Alzheimer's disease (Smith, M.A. et al. (1998) Free Rad. Biol. Med. 25:898-902).

Acyl-CoA thioesterase is another member of the carboxylesterase family (Alexson, S.E. et al. (1993) Eur. J. Biochem. 214:719-727). Evidence suggests that acyl-CoA thioesterase has a regulatory role in steroidogenic tissues (Finkielstein, C. et al. (1998) Eur. J. Biochem. 256:60-66).

The alpha/beta hydrolase protein fold is common to several hydrolases of diverse phylogenetic origin and catalytic function. Enzymes with the alpha/beta hydrolase fold have a common core structure consisting of eight beta-sheets connected by alpha-helices. The most conserved structural feature of this fold is the loops of the nucleophile-histidine-acid catalytic triad. The histidine in the catalytic triad is completely conserved, while the nucleophile and acid loops accommodate more than one type of amino acid (Ollis, D.L. et al. (1992) Protein Eng. 5:197-211).

Sulfatases are members of a highly conserved gene family that share extensive sequence homology and a high degree of structural similarity. Sulfatases catalyze the cleavage of sulfate esters. To perform this function, sulfatases undergo a unique post-translational modification in the endoplasmic reticulum that involves the oxidation of a conserved cysteine residue. A human disorder called multiple sulfatase deficiency is due to a defect in this post-translational modification step, leading to inactive sulfatases (Recksiek, M. et al. (1998) J. Biol. Chem. 273:6096-6103).

Phosphohydrolases are enzymes that hydrolyze phosphate esters. Some phosphohydrolases contain a mutT domain signature sequence. MutT is a protein involved in the GO system responsible for removing an oxidatively damaged form of guanine from DNA. A region of about 40 amino acid residues, found in the N-terminus of mutT, is also found in other proteins, including some phosphohydrolases (PROSITE PDOC00695).

Serine hydrolases are a large functional class of hydrolytic enzymes that contain a serine residue in their active site. This class of enzymes contains proteinases, esterases, and lipases which hydrolyze a variety of substrates and, therefore, have different biological roles. Proteins in this superfamily can be further grouped into subfamilies based on substrate specificity or amino acid similarities (Puente, X.S. and C. Lopez-Otin (1995) J. Biol. Chem. 270:12926-12932).

Neuropathy target esterase (NTE) is an integral membrane protein present in all neurons and in some non-neural-cell types of vertebrates. NTE is involved in a cell-signaling pathway controlling interactions between neurons and accessory glial cells in the developing nervous system. NTE has serine esterase activity and efficiently catalyses the hydrolysis of phenyl valerate (PV) *in vitro*, but its physiological substrate is unknown. NTE is not related to either the major serine esterase family, which includes acetylcholinesterase, nor to any other known serine hydrolases. NTE contains at least

two functional domains: an N-terminal putative regulatory domain and a C-terminal effector domain which contains the esterase activity and is, in part, conserved in proteins found in bacteria, yeast, nematodes and insects. NTE's effector domain contains three predicted transmembrane segments, and the active-site serine residue lies at the center of one of these segments. The isolated recombinant domain shows PV hydrolase activity only when incorporated into phospholipid liposomes. NTE's esterase activity is largely redundant in adult vertebrates, but organophosphates which react with NTE *in vivo* initiate unknown events which lead to a neuropathy with degeneration of long axons. These neuropathic organophosphates leave a negatively charged group covalently attached to the active-site serine residue, which causes a toxic gain of function in NTE (Glynn, P. (1999) Biochem. J. 344:625-631). Further, the *Drosophila* neurodegeneration gene swiss-cheese encodes a neuronal protein involved in glia-neuron interaction and is homologous to the above human NTE (Moser, M. et al. (2000) Mech. Dev. 90:279-282).

Chitinases are chitin-degrading enzymes present in a variety of organisms and participate in processes including cell wall remodeling, defense and catabolism. Chitinase activity has been found in human serum, leukocytes, granulocytes, and in association with fertilized oocytes in mammals (Escott, G.M. (1995) Infect. Immunol. 63:4770-4773; DeSouza, M.M. (1995) Endrocrinology 136:2485-2496). Glycolytic and proteolytic molecules in humans are associated with tissue damage in lung diseases and with increased tumorigenicity and metastatic potential of cancers (Mulligan, M.S. (1993) Proc. Natl. Acad. Sci. 90:11523-11527; Matrisian, L.M. (1991) Am. J. Med. Sci. 302:157-162; Witty, J.P. (1994) Cancer Res. 54:4805-4812). The discovery of a human enzyme with chitinolytic activity is noteworthy given the lack of endogenous chitin in the human body (Raghavan, N. (1994) Infect. Immun. 62:1901-1908). However, there is a group of mammalian proteins that share homology with chitinases from various non-mammalian organisms, such as bacteria, fungi, plants, and insects. The members of this family differ in their ability to hydrolyze chitin or chitin-like substrates. Some of the mammalian members of the family, such as a bovine whey chitotriosidase and human cartilage proteins which do not demonstrate specific chitinolytic activity, are expressed in association with tissue remodeling events (Rejman, J.J. (1988) Biochem. Biophys. Res. Commun. 150:329-334, Nyirkos, P. (1990) Biochem. J. 268:265-268). Elevated levels of human cartilage proteins have been reported in the synovial fluid and cartilage of patients with rheumatoid arthritis, a disease which produces a severe degradation of the cartilage and a proliferation of the synovial membrane in the affected joints (Hakala, B.E. (1993) J. Biol. Chem. 268:25803-25810).

A small subclass of hydrolases acting on ether bonds includes the thioether hydrolases. *S*-adenosyl-L-homocysteine hydrolase, also known as AdoHcyase or SAHH (PROSITE PDOC00603;

EC 3.3.1.1), is a thioether hydrolase first described in rat liver extracts as the activity responsible for the reversible hydrolysis of *S*-adenosyl-L-homocysteine (AdoHcy) to adenosine and homocysteine (Sganga, M.W. et al. (1992) PNAS 89:6328-6332). SAHH is a cytosolic enzyme that has been found in all cells that have been tested, with the exception of *Escherichia coli* and certain related bacteria (Walker, R.D. et al. (1975) Can. J. Biochem. 53:312-319; Shimizu, S. et al. (1988) FEMS Microbiol. Lett. 51:177-180; Shimizu, S. et al. (1984) Eur. J. Biochem. 141:385-392). SAHH activity is dependent on NAD^+ as a cofactor. Deficiency of SAHH is associated with hypermethioninemia (Online Mendelian Inheritance in Man (OMIM) #180960 Hypermethioninemia), a pathologic condition characterized by neonatal cholestasis, failure to thrive, mental and motor retardation, facial dysmorphism with abnormal hair and teeth, and myocaridopathy (Labrune, P. et al. (1990) J. Pediat. 117:220-226).

Another subclass of hydrolases includes those enzymes which act on carbon-nitrogen (C-N) bonds other than peptide bonds. To this subclass belong those enzymes hydrolyzing amides, amidines, and other C-N bonds. This subclass is further subdivided on the basis of substrate specificity such as linear amides, cyclic amides, linear amidines, cyclic amidines, nitriles and other compounds. A hydrolase belonging to the sub-subclass of enzymes acting on the cyclic amidines is adenosine deaminase (ADA). ADA catalyzes the breakdown of adenosine to inosine. ADA is present in many mammalian tissues, including placenta, muscle, lung, stomach, digestive diverticulum, spleen, erythrocytes, thymus, seminal plasma, thyroid, T-cells, bone marrow stem cells, and liver. A subclass of ADAs, ADAR, act on RNA and are classified as RNA editases. An ADAR from *Drosophila*, dADAR, expressed in the developing nervous system, may act on para voltage-gated Na+ channel transcripts in the central nervous system (Palladino, M.J. et al. (2000) RNA 6:1004-1018). ADA deficiency causes profound lymphopenia with severe combined immunodeficiency (SCID). Cells from patients with ADA deficiency contain low, sometimes undetectable, amounts of ADA catalytic activity and ADA protein. ADA deficiency stems from genetic mutations in the ADA gene (Hershfield, M.S. (1998) Semin. Hematol. 4:291-298). Metabolic consequences of ADA deficiency are associated with defects in alveogenesis, pulmonary inflammation, and airway obstruction (Blackburn, M.R. et al. (2000) J. Exp. Med. 192:159-170).

Pancreatic ribonucleases (RNase) are pyrimidine-specific endonucleases found in high quantity in the pancreas of certain mammalian taxa and of some reptiles (Beintema, J.J. et al (1988) Prog. Biophys. Mol. Biol. 51:165-192). Proteins in the mammalian pancreatic RNase superfamily are noncytosolic endonucleases that degrade RNA through a two-step transphosphorolytic-hydrolytic reaction (Beintema, J.J. et al. (1986) Mol. Biol. Evol. 3:262-275). Specifically, the enzymes are

involved in endonucleolytic cleavage of 3'-phosphomononucleotides and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates. Ribonucleases can unwind the DNA helix by complexing with single-stranded DNA; the complex arises by an extended multi-site cation-anion interaction between lysine and arginine residues of the enzyme and phosphate groups of the nucleotides. Some of the enzymes belonging to this family appear to play a purely digestive role, whereas others exhibit potent and unusual biological activities (D'Alessio, G. (1993) Trends Cell Biol. 3:106-109). Proteins belonging to the pancreatic RNase family include: bovine seminal vesicle and brain ribonucleases; kidney non-secretory ribonucleases (Beintema, J.J. et al (1986) FEBS Lett. 194:338-343); liver-type ribonucleases (Rosenberg, H.F. et al. (1989) PNAS U.S.A. 86:4460-4464); angiogenin, which induces vascularisation of normal and malignant tissues; eosinophil cationic protein (Hofsteenge, J. et al. (1989) Biochemistry 28:9806-9813), a cytotoxin and helminthotoxin with ribonuclease activity; and frog liver ribonuclease and frog sialic acid-binding lectin. The sequences of pancreatic RNases contain 4 conserved disulfide bonds and 3 amino acid residues involved in the catalytic activity.

ADP-ribosylation is a reversible post-translational protein modification in which an ADP-ribose moiety is transferred from β-NAD to a target amino acid such as arginine or cysteine. ADP-ribosylarginine hydrolases regenerate arginine by removing ADP-ribose from the protein, completing the ADP-ribosylation cycle (Moss, J. et al. (1997) Adv. Exp. Med. Biol. 419:25-33). ADP-ribosylation is a well-known reaction among bacterial toxins. Cholera toxin, for example, disrupts the adenylyl cyclase system by ADP-ribosylating the α-subunit of the stimulatory G-protein, causing an increase in intracellular cAMP (Moss, J. and M. Vaughan (Eds) (1990) ADP-ribosylating Toxins and G-Proteins: Insights into Signal Transduction, American Society for Microbiology, Washington, D.C.). ADP-ribosylation may also have a regulatory function in eukaryotes, affecting such processes as cytoskeletal assembly (Zhou, H. et al. (1996) Arch. Biochem. Biophys. 334:214-222) and cell proliferation in cytotoxic T-cells (Wang, J. et al. (1996) J. Immunol. 156:2819-2827).

Nucleotidases catalyze the formation of free nucleosides from nucleotides. The cytosolic nucleotidase cN-I (5' nucleotidase-I) cloned from pigeon heart catalyzes the formation of adenosine from AMP generated during ATP hydrolysis (Sala-Newby, G.B. et al. (1999) J. Biol. Chem. 274:17789-17793). Increased adenosine concentration is thought to be a signal of metabolic stress, and adenosine receptors mediate effects including vasodilation, decreased stimulatory neuron firing and ischemic preconditioning in the heart (Schrader, J. (1990) Circulation 81:389-391; Rubino, A. et al. (1992) Eur. J. Pharmacol. 220:95-98; de Jong, J.W. et al. (2000) Pharmacol. Ther. 87:141-149). Deficiency of pyrimidine 5'-nucleotidase can result in hereditary hemolytic anemia (OMIM #266120).

The lysozyme c superfamily consists of conventional lysozymes c, calcium-binding lysozymes c, and α-lactalbumin (Prager, E.M. and P. Jolles (1996) EXS 75:9-31). The proteins in this superfamily have 35-40% sequence homology and share a common three-dimensional fold, but can have different functions. Lysozymes c are ubiquitous in a variety of tissues and secretions and can lyse the cell walls of certain bacteria (McKenzie, H.A. (1996) EXS 75:365-409). Alpha-lactalbumin is a metallo-protein that binds calcium and participates in the synthesis of lactose (Iyer, L.K. and P.K. Qasba (1999) Protein Eng. 12:129-139). Alpha-lactalbumin occurs in mammalian milk and colostrum (McKenzie, *supra*).

Lysozymes catalyze the hydrolysis of certain mucopolysaccharides of bacterial cell walls, specifically, the beta (1-4) glycosidic linkages between N-acetylmuramic acid and N-acetylglucosamine, and cause bacterial lysis. Lysozymes occur in diverse organisms including viruses, birds, and mammals. In humans, lysozymes are found in spleen, lung, kidney, white blood cells, plasma, saliva, milk, tears, and cartilage (OMIM #153450 Lysozyme; Weaver, L.H. et al. (1985) J. Mol. Biol. 184:739-741). Lysozyme c functions in ruminants as a digestive enzyme, releasing proteins from ingested bacterial cells, and may perform the same function in human newborns (Braun, O.H. et al. (1995) Klin. Pediatr. 207:4-7).

The two known forms of lysozymes, chicken-type and goose-type, were originally isolated from chicken and goose egg white, respectively. Chicken-type and goose-type lysozymes have similar three-dimensional structures, but different amino acid sequences (Nakano, T. and T. Graf (1991) Biochim. Biophys. Acta 1090:273-276). In chickens, both forms of lysozyme are found in neutrophil granulocytes (heterophils), but only chicken-type lysozyme is found in egg white. Generally, chicken-type lysozyme mRNA is found in both adherent monocytes and macrophages and nonadherent promyelocytes and granulocytes as well as in cells of the bone marrow, spleen, bursa, and oviduct. Goose-type lysozyme mRNA is found in non-adherent cells of the bone marrow and lung. Several isozymes have been found in rabbits, including leukocytic, gastrointestinal, and possibly lymphoepithelial forms (OMIM #153450, *supra*; Nakano and Graf, *supra*; and GenBank GI 1310929). A human lysozyme gene encoding a protein similar to chicken-type lysozyme has been cloned (Yoshimura, K. et al. (1988) Biochem. Biophys. Res. Commun. 150:794-801). A consensus motif featuring regularly spaced cysteine residues has been derived from the lysozyme C enzymes of various species (PROSITE PS00128). Lysozyme C shares about 40% amino acid sequence identity with α-lactalbumin.

Lysozymes have several disease associations. Lysozymuria is observed in diabetic nephropathy (Shima, M. et al. (1986) Clin. Chem. 32:1818-1822), endemic nephropathy (Bruckner, I.

et al. (1978) Med. Interne. 16:117-125), urinary tract infections (Heidegger, H. (1990) Minerva Ginecol. 42:243-250), and acute monocytic leukemia (Shaw, M.T. (1978) Am. J. Hematol. 4:97-103). Nakano and Graf (*supra*) suggested a role for lysozyme in host defense systems. Older rabbits with an inherited lysozyme deficiency show increased susceptibility to infections, such as subcutaneous abscesses (OMIM #153450, *supra*). Human lysozyme gene mutations cause hereditary systemic amyloidosis, a rare autosomal dominant disease in which amyloid deposits form in the viscera, including the kidney, adrenal glands, spleen, and liver. This disease is usually fatal by the fifth decade. The amyloid deposits contain variant forms of lysozyme. Renal amyloidosis is the most common and potentially the most serious form of organ involvement (Pepys, M.B. et al. (1993) Nature 362:553-557; OMIM #105200 Familial Visceral Amyloidosis; Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Company, Philadelphia PA, pp. 231-238). Increased levels of lysozyme and lactate have been observed in the cerebrospinal fluid of patients with bacterial meningitis (Ponka, A. et al. (1983) Infection 11:129-131). Acute monocytic leukemia is characterized by massive lysozymuria (Den Tandt, W.R. (1988) Int. J. Biochem. 20:713-719).

Lyases

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O, or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, p.620). Under the International Classification of Enzymes (Webb, E. C. (1992) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, San Diego CA), lyases form a distinct class designated by the numeral 4 in the first digit of the enzyme number (i.e., EC 4.x.x.x).

Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group. The group of C-C lyases includes carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases, and other lyases. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides, and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases, amine-lyases (deaminases), and other lyases. Lyases are critical components of cellular biochemistry, with roles in metabolic energy production, including fatty acid metabolism and the tricarboxylic acid cycle, as well as other diverse enzymatic processes.

One important family of lyases are the carbonic anhydrases (CA), also called carbonate dehydratases, which catalyze the hydration of carbon dioxide in the reaction $H_2O + CO_2 \rightleftharpoons HCO_3^- +$

H^+. CA accelerates this reaction by a factor of over 10^6 by virtue of a zinc ion located in a deep cleft about 15Å below the protein's surface and co-ordinated to the imidazole groups of three His residues. Water bound to the zinc ion is rapidly converted to HCO_3^-.

Eight enzymatic and evolutionarily related forms of carbonic anhydrase are currently known to exist in humans: three cytosolic isozymes (CAI, CAII, and CAIII), two membrane-bound forms (CAIV and CAVII), a mitochondrial form (CAV), a secreted salivary form (CAVI) and a yet uncharacterized isozyme (PROSITE PDOC00146 Eukaryotic-type carbonic anhydrases signature). Though the isoenzymes CAI, CAII, and bovine CAIII have similar secondary structures and polypeptide-chain folds, CAI has 6 tryptophans, CAII has 7 and CAIII has 8 (Boren, K. et al. (1996) Protein Sci. 5:2479-2484). CAII is the predominant CA isoenzyme in the brain of mammals.

CAs participate in a variety of physiological processes that involve pH regulation, CO_2 and HCO_3^- transport, ion transport, and water and electrolyte balance. For example, CAII contributes to H^+ secretion by gastric parietal cells, by renal tubular cells, and by osteoclasts that secrete H^+ to acidify the bone-resorbing compartment. In addition, CAII promotes HCO_3^- secretion by pancreatic duct cells, cilary body epithelium, choroid plexus, salivary gland acinar cells, and distal colonal epithelium, thus playing a role in the production of pancreatic juice, aqueous humor, cerebrospinal fluid, and saliva, and contributing to electrolyte and water balance. CAII also promotes CO_2 exchange in proximal tubules in the kidney, in erythrocytes, and in lung. CAIV has roles in several tissues: it facilitates HCO_3^- reabsorption in the kidney; promotes CO_2 flux in tissues including brain, skeletal muscle, and heart muscle; and promotes CO_2 exchange from the blood to the alveoli in the lung. CAVI probably plays a role in pH regulation in saliva, along with CAII, and may have a protective effect in the esophagus and stomach. Mitochondrial CAV appears to play important roles in gluconeogenesis and ureagenesis, based on the effects of CA inhibitors on these pathways. (Sly, W.S. and P.Y. Hu (1995) Ann. Rev. Biochem. 64:375-401.)

A number of disease states are marked by variations in CA activity. Mutations in CAII which lead to CAII deficiency are the cause of osteopetrosis with renal tubular acidosis (OMIM #259730 Osteopetrosis with Renal Tubular Acidosis). The concentration of CAII in the cerebrospinal fluid (CSF) appears to mark disease activity in patients with brain damage. High CA concentrations have been observed in patients with brain infarction. Patients with transient ischemic attack, multiple sclerosis, or epilepsy usually have CAII concentrations in the normal range, but higher CAII levels have been observed in the CSF of those with central nervous system infection, dementia, or trigeminal neuralgia (Parkkila, A.K. et al. (1997) Eur. J. Clin. Invest. 27:392-397). Colonic adenomas and adenocarcinomas have been observed to fail to stain for CA, whereas non-neoplastic controls showed

CAI and CAII in the cytoplasm of the columnar cells lining the upper half of colonic crypts. The neoplasms show staining patterns similar to less mature cells lining the base of normal crypts (Gramlich T.L. et al. (1990) Arch. Pathol. Lab. Med. 114:415-419).

Therapeutic interventions in a number of diseases involve altering CA activity. CA inhibitors such as acetazolamide are used in the treatment of glaucoma (Stewart, W.C. (1999) Curr. Opin. Opthamol. 10:99-108), essential tremor and Parkinson's disease (Uitti, R.J. (1998) Geriatrics 53:46-48, 53-57), intermittent ataxia (Singhvi, J.P. et al. (2000) Neurology India 48:78-80), and altitude related illnesses (Klocke, D.L. et al. (1998) Mayo Clin. Proc. 73:988-992).

CA activity can be particularly useful as an indicator of long-term disease conditions, since the enzyme reacts relatively slowly to physiological changes. CAI and zinc concentrations have been observed to decrease in hyperthyroid Graves' disease (Yoshida, K. (1996) Tohoku J. Exp. Med. 178:345-356) and glycosylated CAI is observed in diabetes mellitus (Kondo, T. et al. (1987) Clin. Chim. Acta 166:227-236). A positive correlation has been observed between CAI and CAII reactivity and endometriosis (Brinton, D.A. et al. (1996) Ann. Clin. Lab. Sci. 26:409-420; D'Cruz , O.J. et al. (1996) Fertil. Steril. 66:547-556).

Another important member of the lyase family is ornithine decarboxylase (ODC), the initial rate-limiting enzyme in polyamine biosynthesis. ODC catalyses the transformation of ornithine into putrescine in the reaction L-ornithine $\rightleftharpoons$ putrescine + CO_2. Polyamines, which include putrescine and the subsequent metabolic pathway products spermidine and spermine, are ubiquitous cell components essential for DNA synthesis, cell differentiation, and proliferation. Thus the polyamines play a key role in tumor proliferation (Medina, M.A. et al. (1999) Biochem. Pharmacol. 57:1341-1344).

ODC is a pyridoxal-5'-phosphate (PLP)-dependent enzyme which is active as a homodimer. Conserved residues include those at the PLP binding site and a stretch of glycine residues thought to be part of a substrate binding region (PROSITE PDOC00685 Orn/DAP/Arg decarboxylase family 2 signatures). Mammalian ODCs also contain PEST regions, sequence fragments enriched in proline, glutamic acid, serine, and threonine residues that act as signals for intracellular degradation (Medina et al., *supra*).

Many chemical carcinogens and tumor promoters increase ODC levels and activity. Several known oncogenes may increase ODC levels by enhancing transcription of the ODC gene, and ODC itself may act as an oncogene when expressed at very high levels. A high level of ODC is found in a number of precancerous conditions, and elevation of ODC levels has been used as part of a screen for tumor-promoting compounds (Pegg, A.E. et al. (1995) J. Cell. Biochem. Suppl. 22:132-138).

Inhibitors of ODC have been used to treat tumors in animal models and human clinical trials,

and have been shown to reduce development of tumors of the bladder, brain, esophagus, gastrointestinal tract, lung, oral cavity, mammary gland, stomach, skin and trachea (Pegg et al., *supra*; McCann, P.P. and A.E. Pegg (1992) Pharmac. Ther. 54:195-215). ODC also shows promise as a target for chemoprevention (Pegg et al., *supra*). ODC inhibitors have also been used to treat infections by African trypanosomes, malaria, and *Pneumocystis carinii*, and are potentially useful for treatment of autoimmune diseases such as lupus and rheumatoid arthritis (McCann and Pegg, *supra*).

Another family of pyridoxal-dependent decarboxylases are the group II decarboxylases. This family includes glutamate decarboxylase (GAD) which catalyzes the decarboxylation of glutamate into the neurotransmitter GABA; histidine decarboxylase (HDC), which catalyzes the decarboxylation of histidine to histamine; aromatic-L-amino-acid decarboxylase (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase, which catalyzes the decarboxylation of tryptophan to tryptamine and also acts on 5-hydroxy-tryptophan and dihydroxyphenylalanine (L-dopa); and cysteine sulfinic acid decarboxylase (CSD), the rate-limiting enzyme in the synthesis of taurine from cysteine (PROSITE PDOC00329 DDC/GAD/HDC/TyrDC pyridoxal-phosphate attachment site). Taurine is an abundant sulfonic amino acid in brain and is thought to act as an osmoregulator in brain cells (Bitoun, M. and M. Tappaz (2000) J. Neurochem. 75:919-924).

<u>Isomerases</u>

Isomerases are a class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. This class includes racemases and epimerases, cis-trans-isomerases, intramolecular oxidoreductases, intramolecular transferases (mutases) and intramolecular lyases. Isomerases are critical components of cellular biochemistry with roles in metabolic energy production including glycolysis, as well as other diverse enzymatic processes (Stryer, *supra*, pp.483-507).

Racemases are a subset of isomerases that catalyze inversion of a molecule's configuration around the asymmetric carbon atom in a substrate having a single center of asymmetry, thereby interconverting two racemers. Epimerases are another subset of isomerases that catalyze inversion of configuration around an asymmetric carbon atom in a substrate with more than one center of symmetry, thereby interconverting two epimers. Racemases and epimerases can act on amino acids and derivatives, hydroxy acids and derivatives, and carbohydrates and derivatives. The interconversion of UDP-galactose and UDP-glucose is catalyzed by UDP-galactose-4'-epimerase. Proper regulation and function of this epimerase is essential to the synthesis of glycoproteins and glycolipids. Elevated blood galactose levels have been correlated with UDP-galactose-4'-epimerase

deficiency in screening programs of infants (Gitzelmann, R. (1972) Helv. Paediat. Acta 27:125-130).

Correct folding of newly synthesized proteins is assisted by molecular chaperones and folding catalysts, two unrelated groups of helper molecules. Chaperones suppress non-productive side reactions by stoichiometric binding to folding intermediates, whereas folding enzymes catalyze some of the multiple folding steps that enable proteins to attain their final functional configurations (Kern, G. et al. (1994) FEBS Lett. 348:145-148). One class of folding enzymes, the peptidyl prolyl *cis-trans* isomerases (PPIases), isomerizes certain proline imidic bonds in what is considered to be a rate limiting step in protein maturation and export. PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. There are three evolutionarily unrelated families of PPIases: the cyclophilins, the FK506 binding proteins, and the newly characterized parvulin family (Rahfeld, J.U. et al. (1994) FEBS Lett. 352:180-184).

The cyclophilins (CyP) were originally identified as major receptors for the immunosuppressive drug cyclosporin A (CsA), an inhibitor of T-cell activation (Handschumacher, R.E. et al. (1984) Science 226:544-547; Harding, M.W. et al. (1986) J. Biol. Chem. 261:8547-8555). Thus, the peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. Subsequent work demonstrated that CyP's isomerase activity is essential for correct protein folding and/or protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsp70 complex that binds steroid receptors. The mammalian CyP (CypA) has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyP in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214; Hunter, T. (1998) Cell 92:141-143; and Leverson, J.D. and S.A. Ness (1998) Mol. Cell. 1:203-211).

One of the major rate limiting steps in protein folding is the thiol:disulfide exchange that is necessary for correct protein assembly. Although incubation of reduced, unfolded proteins in buffers with defined ratios of oxidized and reduced thiols can lead to native conformation, the rate of folding is slow and the attainment of native conformation decreases proportionately with the size and number of cysteines in the protein. Certain cellular compartments such as the endoplasmic reticulum of eukaryotes and the periplasmic space of prokaryotes are maintained in a more oxidized state than the

surrounding cytosol. Correct disulfide formation can occur in these compartments, but at a rate that is insufficient for normal cell processes and inadequate for synthesizing secreted proteins. The protein disulfide isomerases, thioredoxins and glutaredoxins are able to catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges (Loferer, H. (1995) J. Biol. Chem. 270:26178-26183).

Each of these proteins has somewhat different functions, but all belong to a group of disulfide-containing redox proteins that contain a conserved active-site sequence and are ubiquitously distributed in eukaryotes and prokaryotes. Protein disulfide isomerases are found in the endoplasmic reticulum of eukaryotes and in the periplasmic space of prokaryotes. They function by exchanging their own disulfide for a thiol in a folding peptide chain. In contrast, the reduced thioredoxins and glutaredoxins are generally found in the cytoplasm and function by directly reducing disulfides in the substrate proteins.

Oxidoreductases can be isomerases as well. Oxidoreductases catalyze the reversible transfer of electrons from a substrate that becomes oxidized to a substrate that becomes reduced. This class of enzymes includes dehydrogenases, hydroxylases, oxidases, oxygenases, peroxidases, and reductases. Proper maintenance of oxidoreductase levels is physiologically important. For example, genetically-linked deficiencies in lipoamide dehydrogenase can result in lactic acidosis (Robinson, B. H. et al. (1977) Pediat. Res. 11:1198-1202).

Another subgroup of isomerases are the transferases (or mutases). Transferases transfer a chemical group from one compound (the donor) to another compound (the acceptor). The types of groups transferred by these enzymes include acyl groups, amino groups, phosphate groups (phosphotransferases or phosphomutases), and others. The transferase carnitine palmitoyltransferase is an important component of fatty acid metabolism. Genetically-linked deficiencies in this transferase can lead to myopathy (Scriver, C. et al. (1995) The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill, New York NY, pp.1501-1533).

Yet another subgroup of isomerases are the topoisomersases. Topoisomerases are enzymes that affect the topological state of DNA. For example, defects in topoisomerases or their regulation can affect normal physiology. Reduced levels of topoisomerase II have been correlated with some of the DNA processing defects associated with the disorder ataxia-telangiectasia (Singh, S.P. et al. (1988) Nucleic Acids Res. 16:3919-3929).

Ligases

Ligases catalyze the formation of a bond between two substrate molecules. The process

involves the hydrolysis of a pyrophosphate bond in ATP or a similar energy donor. Ligases are classified based on the nature of the type of bond they form, which can include carbon-oxygen, carbon-sulfur, carbon-nitrogen, carbon-carbon and phosphoric ester bonds.

Ligases forming carbon-oxygen bonds include the aminoacyl-transfer RNA (tRNA) synthetases which are important RNA-associated enzymes with roles in translation. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for the activation and correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes can be divided into two structural classes, and each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding "Rossman fold". Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet motif, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and S. Cusack, (1995) J. Mol. Evol. 40:519-530). Autoantibodies against aminoacyl-tRNAs are generated by patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Ligases forming carbon-sulfur bonds (acid-thiol ligases) mediate a large number of cellular biosynthetic intermediary metabolism processes involving intermolecular transfer of carbon atom-containing substrates (carbon substrates). Examples of such reactions include the tricarboxylic acid cycle, synthesis of fatty acids and long-chain phospholipids, synthesis of alcohols and aldehydes, synthesis of intermediary metabolites, and reactions involved in the amino acid degradation pathways. Some of these reactions require input of energy, usually in the form of conversion of ATP to either ADP or AMP and pyrophosphate.

In many cases, a carbon substrate is derived from a small molecule containing at least two carbon atoms. The carbon substrate is often covalently bound to a larger molecule which acts as a carbon substrate carrier molecule within the cell. In the biosynthetic mechanisms described above, the carrier molecule is coenzyme A. Coenzyme A (CoA) is structurally related to derivatives of the nucleotide ADP and consists of 4'-phosphopantetheine linked via a phosphodiester bond to the alpha phosphate group of adenosine 3',5'-bisphosphate. The terminal thiol group of 4'-phosphopantetheine acts as the site for carbon substrate bond formation. The predominant carbon substrates which utilize CoA as a carrier molecule during biosynthesis and intermediary metabolism in the cell are acetyl,

succinyl, and propionyl moieties, collectively referred to as acyl groups. Other carbon substrates include enoyl lipid, which acts as a fatty acid oxidation intermediate, and carnitine, which acts as an acetyl-CoA flux regulator/mitochondrial acyl group transfer protein. Acyl-CoA and acetyl-CoA are synthesized in the cell by acyl-CoA synthetase and acetyl-CoA synthetase, respectively.

Activation of fatty acids is mediated by at least three forms of acyl-CoA synthetase activity: i) acetyl-CoA synthetase, which activates acetate and several other low molecular weight carboxylic acids and is found in muscle mitochondria and the cytosol of other tissues; ii) medium-chain acyl-CoA synthetase, which activates fatty acids containing between four and eleven carbon atoms (predominantly from dietary sources), and is present only in liver mitochondria; and iii) acyl CoA synthetase, which is specific for long chain fatty acids with between six and twenty carbon atoms, and is found in microsomes and the mitochondria. Proteins associated with acyl-CoA synthetase activity have been identified from many sources including bacteria, yeast, plants, mouse, and man. The activity of acyl-CoA synthetase may be modulated by phosphorylation of the enzyme by cAMP-dependent protein kinase.

Ligases forming carbon-nitrogen bonds include amide synthases such as glutamine synthetase (glutamate-ammonia ligase) that catalyzes the amination of glutamic acid to glutamine by ammonia using the energy of ATP hydrolysis. Glutamine is the primary source for the amino group in various amide transfer reactions involved in *de novo* pyrimidine nucleotide synthesis and in purine and pyrimidine ribonucleotide interconversions. Overexpression of glutamine synthetase has been observed in primary liver cancer (Christa, L. et al. (1994) Gastroent. 106:1312-1320).

Acid-amino-acid ligases (peptide synthases) are represented by the ubiquitin conjugating enzymes which are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin (Ub), a small heat stable protein. Ub is first activated by a ubiquitin-activating enzyme (E1), and then transferred to one of several Ub-conjugating enzymes (E2). E2 then links the Ub molecule through its C-terminal glycine to an internal lysine (acceptor lysine) of a target protein. The ubiquitinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and ubiquitin is released for reutilization by ubiquitin protease. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, A.

(1994) Cell 79:13-21).

Cyclo-ligases and other carbon-nitrogen ligases comprise various enzymes and enzyme complexes that participate in the *de novo* pathways of purine and pyrimidine biosynthesis. Because these pathways are critical to the synthesis of nucleotides for replication of both RNA and DNA, many of these enzymes have been the targets of clinical agents for the treatment of cell proliferative disorders such as cancer and infectious diseases.

Purine biosynthesis occurs *de novo* from the amino acids glycine and glutamine, and other small molecules. Three of the key reactions in this process are catalyzed by a trifunctional enzyme composed of glycinamide-ribonucleotide synthetase (GARS), aminoimidazole ribonucleotide synthetase (AIRS), and glycinamide ribonucleotide transformylase (GART). Together these three enzymes combine ribosylamine phosphate with glycine to yield phosphoribosyl aminoimidazole, a precursor to both adenylate and guanylate nucleotides. This trifunctional protein has been implicated in the pathology of Downs syndrome (Aimi, J. et al. (1990) Nucleic Acid Res. 18:6665-6672). Adenylosuccinate synthetase catalyzes a later step in purine biosynthesis that converts inosinic acid to adenylosuccinate, a key step on the path to ATP synthesis. This enzyme is also similar to another carbon-nitrogen ligase, argininosuccinate synthetase, that catalyzes a similar reaction in the urea cycle (Powell, S.M. et al. (1992) FEBS Lett. 303:4-10).

Adenylosuccinate synthetase, adenylosuccinate lyase, and AMP deaminase may be considered as a functional unit, the purine nucleotide cycle. This cycle converts AMP to inosine monophosphate (IMP) and reconverts IMP to AMP via adenylosuccinate, thereby producing NH_3 and forming fumarate from aspartate. In muscle, the purine nucleotide cycle functions, during intense exercise, in the regeneration of ATP by pulling the adenylate kinase reaction in the direction of ATP formation and by providing Krebs cycle intermediates. In kidney, the purine nucleotide cycle accounts for the release of NH_3 under normal acid-base conditions. In brain, the purine nucleotide cycle may contribute to ATP recovery. Adenylosuccinate lyase deficiency provokes psychomotor retardation, often accompanied by autistic features (Van den Berghe, G. et al. (1992) Prog Neurobiol.: 39:547-561). A marked imbalance in the enzymic pattern of purine metabolism is linked with transformation and/or progression in cancer cells. In rat hepatomas the specific activities of the anabolic enzymes, IMP dehydrogenase, GMP synthetase, adenylosuccinate synthetase, adenylosuccinase, AMP deaminase and amidophosphoribosyltransferase, increased to 13.5-, 3.7-, 3.1-, 1.8-, 5.5- and 2.8-fold, respectively, of those in normal liver (Weber, G. (1983) Clin. Biochem. 16:57-63).

Like the *de novo* biosynthesis of purines, *de novo* synthesis of the pyrimidine nucleotides

uridylate and cytidylate also arises from a common precursor, in this instance the nucleotide orotidylate derived from orotate and phosphoribosyl pyrophosphate (PPRP). Again a trifunctional enzyme comprising three carbon-nitrogen ligases plays a key role in the process. In this case the enzymes aspartate transcarbamylase (ATCase), carbamyl phosphate synthetase II, and dihydroorotase (DHOase) are encoded by a single gene called CAD. Together these three enzymes combine the initial reactants in pyrimidine biosynthesis, glutamine, CO_2, and ATP to form dihydroorotate, the precursor to orotate and orotidylate (Iwahana, H. et al. (1996) Biochem. Biophys. Res. Commun. 219:249-255). Further steps then lead to the synthesis of uridine nucleotides from orotidylate. Cytidine nucleotides are derived from uridine-5'-triphosphate (UTP) by the amidation of UTP using glutamine as the amino donor and the enzyme CTP synthetase. Regulatory mutations in the human CTP synthetase are believed to confer multi-drug resistance to agents widely used in cancer therapy (Yamauchi, M. et al. (1990) EMBO J. 9:2095-2099).

Ligases forming carbon-carbon bonds include the carboxylases acetyl-CoA carboxylase and pyruvate carboxylase. Acetyl-CoA carboxylase catalyzes the carboxylation of acetyl-CoA from CO_2 and H_2O using the energy of ATP hydrolysis. Acetyl-CoA carboxylase is the rate-limiting enzyme in the biogenesis of long-chain fatty acids. Two isoforms of acetyl-CoA carboxylase, types I and types II, are expressed in human in a tissue-specific manner (Ha, J. et al. (1994) Eur. J. Biochem. 219:297-306). Pyruvate carboxylase is a nuclear-encoded mitochondrial enzyme that catalyzes the conversion of pyruvate to oxaloacetate, a key intermediate in the citric acid cycle.

Ligases forming phosphoric ester bonds include the DNA ligases involved in both DNA replication and repair. DNA ligases seal phosphodiester bonds between two adjacent nucleotides in a DNA chain using the energy from ATP hydrolysis to first activate the free 5'-phosphate of one nucleotide and then react it with the 3'-OH group of the adjacent nucleotide. This resealing reaction is used in DNA replication to join small DNA fragments called "Okazaki" fragments that are transiently formed in the process of replicating new DNA, and in DNA repair. DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Bloom's syndrome is an inherited human disease in which individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts et al., *supra*, p. 247).

Pantothenate synthetase (D-pantoate; beta-alanine ligase (AMP-forming); EC 6.3.2.1) is the last enzyme of the pathway of pantothenate (vitamin B(5)) synthesis. It catalyzes the condensation of pantoate with beta-alanine in an ATP-dependent reaction. The enzyme is dimeric, with two well-defined domains per protomer: the N-terminal domain, a Rossmann fold, contains the active site

cavity, with the C-terminal domain forming a hinged lid. The N-terminal domain is structurally very similar to class I aminoacyl-tRNA synthetases and is thus a member of the cytidylyltransferase superfamily (von Delft, F. et al. (2000) Structure (Camb) 9:439-450).

Farnesyl diphosphate synthase (FPPS) is an essential enzyme that is required both for cholesterol synthesis and protein prenylation. The enzyme catalyzes the formation of farnesyl diphosphate from dimethylallyl diphosphate and isopentyl diphosphate. FPPS is inhibited by nitrogen-containing biphosphonates, which can lead to the inhibition of osteoclast-mediated bone resorption by preventing protein prenylation (Dunford, J.E. et al. (2001) J. Pharmacol. Exp. Ther. 296:235-242).

5-aminolevulinate synthase (ALAS; delta-aminolevulinate synthase; EC 2.3.1.37) catalyzes the rate-limiting step in heme biosynthesis in both erythroid and non-erythroid tissues. This enzyme is unique in the heme biosynthetic pathway in being encoded by two genes, the first encoding ALAS1, the non-erythroid specific enzyme which is ubiquitously expressed, and the second encoding ALAS2, which is expressed exclusively in erythroid cells. The genes for ALAS1 and ALAS2 are located, respectively, on chromosome 3 and on the X chromosome. Defects in the gene encoding ALAS2 result in X-linked sideroblastic anemia. Elevated levels of ALAS are seen in acute hepatic porphyrias and can be lowered by zinc mesoporphyrin.

Drug Metabolizing Enzymes (DMEs)

The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics. It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. Advances in pharmacogenomics research, of which DMEs constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W.E. and R.V. Relling (1999) Science 286:487-491). DMEs have broad substrate specificities, unlike antibodies, for example, which are diverse and highly specific. Since DMEs metabolize a wide

variety of molecules, drug interactions may occur at the level of metabolism so that, for example, one compound may induce a DME that affects the metabolism of another compound.

Drug metabolic reactions are categorized as Phase I, which prepare the drug molecule for functioning and further metabolism, and Phase II, which are conjugative. In general, Phase I reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[*a*]pyrene) are metabolized to toxic intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C.D. et al. (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; Katzung, B.G. (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.).

The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and

xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; Graham-Lorence, S. and J.A. Peterson (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence and Peterson, *supra*). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, *supra*; Graham-Lorence and Peterson, *supra*.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and F.J. Gonzalez (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (OMIM #601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed *in vivo* can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to

two cytochrome P450 proteins were found in patients with autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM #240300 Autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM #213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D.C. et al. (1999; FEBS Lett. 462:283-288) identifies a *Candida albicans* cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and A.A. Lurie (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D_2), produced in plant tissues, and cholecalciferol (vitamin D_3), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in

Miller, W.L. and A.A. Portale (2000) Trends Endocrinol. Metab. 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1α,25-dihydroxyvitamin D (1α,25(OH)$_2$D), by the enzyme 25-hydroxyvitamin D 1α-hydroxylase (1α-hydroxylase). Regulation of 1α,25(OH)$_2$D production is primarily at this final step in the synthetic pathway. The activity of 1α-hydroxylase depends upon several physiological factors including the circulating level of the enzyme product (1α,25(OH)$_2$D) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α-hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1α,25(OH)2D production may also be biologically important. The catalysis of 1α,25(OH)$_2$D to 24,25-dihydroxyvitamin D (24,25(OH)$_2$D), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12920-12925; Miller and Portale, *supra*; and references within).

Vitamin D 25-hydroxylase, 1α-hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. et al. (1995) J. Biol. Chem. 270:16766-16774; Miller and Portale, *supra*; and references within).

The active form of vitamin D (1α,25(OH)$_2$D) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1α-hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts. Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and J.E. Zerwekh (1983) J. Clin. Invest. 72:1190-1199; Gamblin, G.T. et al. (1985) J. Clin. Invest. 75:954-960; and Miller and Portale, *supra*).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support

at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F.J. et al. (1996) Biochem. J. 320:267-71). A *Streptomyces griseus* cytochrome P450, CYP104D1, was heterologously expressed in *Escherichia coli* and found to be reduced by the endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-842), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W.D. and R.P. Mason (1988) Arch. Biochem. Biophys. 267:632-639).

Flavin-containing monooxygenase (FMO)

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O_2; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

Isoforms of FMO in mammals include FMO1, FMO2, FMO3, FMO4, and FMO5, which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature). Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. FMOs are more heat labile and less detergent-sensitive than cytochromes P450 *in vitro* though FMO isoforms vary in thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H_2-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

Lysyl oxidase

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the

formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as an N-glycosylated precursor protein of approximately 50 kDa and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electrons to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity have been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and H.M. Kagan (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the *de novo* synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:

$$\text{7,8-dihydrofolate} + \text{NADPH} \rightarrow \text{5,6,7,8-tetrahydrofolate} + \text{NADP}^+$$

The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethroprim and methotrexate. Since an abundance of dTMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-519).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K.M. et al. (1989) J. Biol. Chem. 264:9547-9551). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM #103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-11435).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD^+, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b_1, b_2, b_3, g_1, g_2). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-

cyclohexadiene-1-carboxylate dehydrogenase, *cis*-toluene dihydrodiol dehydrogenase, *cis*-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

Sulfotransferases

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring SO_3^- from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem. 259:13751-13757; OMIM #217800 Macular

dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. Known human and mouse β1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger et al., *supra;* and Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet et al., *supra*). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase (Yuan, Y. et al. (1997) Cell 88:9-11; and Hennet et al., *supra*).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) EMBO J. 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages. As is the case with the β1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and K. Brew (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and K. Brew (1991) J. Biol. Chem. 266:698-703; and Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β1,4-galactosyltransferase, as part of a heterodimer with α-lactalbumin, functions in lactating mammary gland lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur,

B. (1993) Curr. Opin. Cell Biol. 5:854-863; and Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104).

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative stress. The cell surface-localized glycoproteins.are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidase activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-342; Taniguchi, N. and Y. Ikeda (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-278; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-380).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyze the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission; thus a deficiency in kynurenine aminotransferase may be associated with pleotrophic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-*O*-methyltransferase

Catechol-*O*-methyltransferase (COMT) catalyzes the transfer of the methyl group of *S*-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed S_N2-like methylation reaction requires Mg^{++} and is inhibited by Ca^{++}. The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a Mg^{++}-independent manner, followed by the binding of Mg^{++} and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for *in vitro* use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiophetropolone) and for clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimiterol, dobutamine, fenoldopam, apomorphine, and α-methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be usefull in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and S. Kaakkola (1999) Pharmacol. Rev. 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into O_2 and H_2O_2. The rate of dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible

for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70 °C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addition, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organism's survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with *Mycobacterium tuberculosis*, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by *M. tuberculosis* and its expression is upregulated approximately 5-fold in response to oxidative stress. *M. tuberculosis* expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium *M. smegmatis*, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by *M. tuberculosis* than *M. smegmatis*, providing substantial resistance to oxidative stress (Harth, G. and M.A. Horwitz (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases is reduced in prostatic intraepithelial neoplasia and prostate carcinomas, (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphoesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. Phosphotriesterases play a central role in the detoxification of insecticides by mammals. Birds and insects lack PTE, and as a result have reduced tolerance for organophosphorus compounds (Vilanova, E. and M.A. Sogorb (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterase activity varies among individuals and is lower in infants than adults. PTE knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). Phosphotriesterase is also implicated in atherosclerosis and diseases involving lipoprotein metabolism.

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester

phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from *E. coli* has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev. 75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are Ca^{2+}/calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated *in vitro* by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar et al., *supra*). PDE1s may provide useful therapeutic targets for disorders of the central nervous system and the cardiovascular and immune systems, due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry and Higgs, *supra*).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion,

participate in the regulation of aldosterone (Beavo, *supra*), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) Curr. Opin. Chem. Biol. 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) Biochemistry 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) J. Biol. Chem. 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) Bioorg. Med. Chem. Lett. 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry and Higgs, *supra*).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the

phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) Methods 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) Invest. Opthalmol. Vis. Sci. 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) Genomics 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) Proc. Natl. Acad. Sci. USA 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) J. Biol. Chem. 272:16152-16157; Perry and Higgs, *supra*). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, *supra*).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone, IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) Biochem. Biophys. Res. Commun. 246:570-577; Hayashi, M. et al. (1998) Biochem. Biophys. Res. Commun. 250:751-756; Soderling, S.H. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) J. Biol. Chem. 273:15559-15564; Soderling, S.H. et al. (1998) J. Biol. Chem. 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:7071-7076; Fujishige, K. et al. (1999) J. Biol. Chem. 274:18438-18445; Loughney, K. et al (1999) Gene 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal

regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti and Jin, *supra*). A conserved, putative zinc-binding motif has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of a conserved sequence motif (McAllister-Lucas, L.M. et al. (1993) J. Biol. Chem. 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) J. Biol. Chem. 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-K_m cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have undergone clinical evaluation (Perry and Higgs, *supra*; Torphy, T.J. (1998) Am. J. Respir. Crit. Care Med. 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other PDE4 inhibitors have an anti-inflammatory effect. Rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) AIDS 9:1137-1144). Additionally, rolipram suppresses the production of cytokines such as TNF-a and b and interferon g, and thus is effective against encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and multiple sclerosis (Sommer, N. et al. (1995) Nat. Med. 1:244-248; Sasaki, H. et al. (1995) Eur. J. Pharmacol. 282:71-76). Theophylline is a nonspecific PDE inhibitor used in treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF-a

production and may inhibit HIV-1 replication (Angel et al., *supra*).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) Endocrine Rev. 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) Proc. Natl. Acad. Sci. USA 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors can regulate mesangial cell proliferation (Matousovic, K. et al. (1995) J. Clin. Invest. 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) J. Lipid Mediat. Cell Signal. 11:63-79). One cancer treatment involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) Br. J. Cancer 70:786-794).

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450 enzymes and flavin-containing monooxygenases. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (PROSITE PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM #191740 UGT1).

Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). *E. coli* TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, *E. coli* TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in *E. coli*, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., *supra*). For that reason, Naggert et al. (*supra*) proposed that the physiological substrates for *E. coli* TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopanthetheine-fatty acid esters.

Carboxylesterases

Mammalian carboxylesterases are a multigene family expressed in a variety of tissues and cell types. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine superfamily of esterases (B-esterases). Other carboxylesterases include thyroglobulin, thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide-groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilicaine, and isocarboxazide. Carboxylesterases are also important for the conversion of prodrugs to free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) Annu. Rev. Pharmacol. Toxicol.38:257-288). Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-

surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682).

<u>Squalene epoxidase</u>

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol.

High serum cholesterol levels result in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels results in decreased blood flow and potential necrosis. HMG-CoA reductase is responsible for the first committed step in cholesterol biosynthesis, conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels, but inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates required for other biochemical pathways. Since SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway with cholesterol as the only end product, SE is a better ideal target for the design of anti-hyperlipidemic drugs (Nakamura, Y. et al. (1996) 271:8053-8056).

<u>Epoxide hydrolases</u>

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the α/β hydrolase fold family of enzymes. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced. Examples of epoxide hydrolase reactions include the hydrolysis of some leukotoxin to leukotoxin diol, and isoleukotoxin to isoleukotoxin diol. Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins. Epoxide hydrolases possess a catalytic triad composed of Asp, Asp, and His (Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine, to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. Enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase, 2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase. Enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, maleylacetoacetate isomerase, fumarylacetoacetase and 4-hydroxyphenylacetate. Additional enzymes associated with tyrosine metabolism in different organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) Nucleic Acids Res. 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) J. Microbiol. Meth. 25:91-93; and Schmidt, M. (1996) Amer. Soc. Microbiol. News 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 (HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) J. Biol. Chem. 272:24426-24432).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the

expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Cancer

Breast cancer

Breast cancer is the most frequently diagnosed type of cancer in American women and the second most frequent cause of cancer death. The lifetime risk of an American woman developing breast cancer is 1 in 8, and one-third of women diagnosed with breast cancer die of the disease. A number of risk factors have been identified, including hormonal and genetic factors. One genetic defect associated with breast cancer results in a loss of heterozygosity (LOH) at multiple loci such as p53, Rb, BRCA1, and BRCA2. Another genetic defect is gene amplification involving genes such as c-myc and c-erbB2 (Her2-neu gene). Steroid and growth factor pathways are also altered in breast cancer, notably the estrogen, progesterone, and epidermal growth factor (EGF) pathways. Breast cancer evolves through a multi-step process whereby premalignant mammary epithelial cells undergo a relatively defined sequence of events leading to tumor formation. An early event in tumor development is ductal hyperplasia. Cells undergoing rapid neoplastic growth gradually progress to invasive carcinoma and become metastatic to the lung, bone, and potentially other organs. Variables that may influence the process of tumor progression and malignant transformation include genetic factors, environmental factors, growth factors, and hormones.

Tumor cells stimulate the formation of stroma that secretes various mediators, such as growth factors, cytokines, and proteases, which are critical for tumor growth. For instance, serum tumor necrosis factor alpha (TNF-α) is increased in the circulation of patients with malignancy. Interferon-gamma (IFN-γ), also known as Type II interferon or immune interferon, is produced primarily by T-lymphocytesand natural killer cells. IFN-γ was originally characterized based on its antiviral characteristics. The protein exhibits antiproliferative, immunoregulatory and proinflammatory activities and is thus important in host defense mechanisms. IFN-γ induces the production of cytokines, upregulates the expression of class I and II MHC antigens, Fc receptor, and leukocyte adhesion molecules. It modulates macrophage effector functions, influences isotype switching and potentiates the secretion of immunoglobulins by B cells. IFN-γ also augments TH1 cell expansion and may be required for TH1 cell differentiation. The IFN-γ receptor has been cloned and characterized, and is

structurally related to the recently cloned IL-10 receptor. It is present on almost all cell types except mature erythrocytes. T-47D is a breast carcinoma cell line isolated from a pleural effusion obtained from a donor with an infiltrating ductal carcinoma of the breast.

Colon cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Lung cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

Osteosarcoma

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy,

treatment involves the use of 3–4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30–40% of patients with non-metastatic disease relapse after therapy. Currently, there is no prognostic factor that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome.

Ovarian cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Prostate cancer

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence *in situ* hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher

frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer Genet. Cytogenet. 124: 20-26).

Tangier Disease

Tangier disease (TD) is a genetic disorder characterized by the near absence of circulating high density lipoprotein (HDL) and the accumulation of cholesterol esters in many tissues, including tonsils, lymph nodes, liver, spleen, thymus, and intestine. Low levels of HDL represent a clear predictor of premature coronary artery disease and homozygous TD correlates with a four- to six-fold increase in cardiovascular disease compared to controls. HDL plays a cardio-protective role in reverse cholesterol transport, the flux of cholesterol from peripheral cells such as tissue macrophages through plasma lipoproteins to the liver. The HDL protein, apolipoprotein A-I, plays a major role in this process, interacting with the cell surface to remove excess cholesterol and phospholipids. This pathway is severely impaired in TD and the defect lies in a specific gene, the ABC1 transporter. This gene is a member of the family of ATP-binding cassette transporters, which utilize ATP hydrolysis to transport a variety of substrates across membranes.

Alzheimer's Disease

Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain

injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

C3A Cells

The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein; iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am. J. Physiol. 272:G408-G416).

Gemfibrozil is a fibric acid antilipemic agent that lowers serum triglycerides and produces favorable changes in lipoproteins. Gemfibrozil is effective in reducing the risk of coronary heart disease in men (Frick, M.H., et al. (1987) New Engl. J. Med; 317:1237-1245). The compound can inhibit peripheral lipolysis and decrease hepatic extraction of free fatty acids, which decreases hepatic triglyceride production. Gemfibrozil also inhibits the synthesis and increases the clearance of apolipoprotein B, a carrier molecule for VLDL. Gemfibrozil has variable effects on LDL cholesterol. Although it causes moderate reductions in patients with type IIa hyperlipoproteinemia, changes in patients with either type IIb or type IV hyperlipoproteinemia are unpredictable. In general, the HMG-CoA reductase inhibitors are more effective than gemfibrozil in reducing LDL cholesterol. At the molecular level gemfibozil may function as a peroxisome proliferator-activated receptor (PPAR) agonist. Gemfibrozil is rapidly and completely absorbed from the GI tract and undergoes enterohepatic recirculation. Gemfibrozil is metabolized by the liver and excreted by the kidneys, mainly as metabolites, one of which possesses pharmacologic activity. Gemfibozil causes peroxisome proliferation and hepatocarcinogenesis in rats, which is a cause for concern generally for fibric acid derivative drugs. In humans, fibric acid derivatives are known to increase the risk of gall bladder

disease although gemfibrozil is better tolerated than other fibrates. The relative safety of gemfibrozil in humans compared to rodent species including rats may be attributed to differences in metabolism and clearance of the compound in different species (Dix, K.J., et al. (1999) Drug Metab. Distrib. 27:138-146; Thomas, B.F., et al. (1999) Drug Metab. Distrib. 27:147-157).

Dendritic cells

Dendritic cells (DCs) are antigen presenting cells that play a crucial role in the initiation of the immune response. DCs can be derived *in vitro* either from CD34+ bone marrow precursors (IDCs) or from peripheral blood monocytic cells (mDCs). *In vivo*, DCs reside in two distinct compartments: the peripheral tissues such as lung, skin, kidney, heart, and intestine; and in secondary lymphoid organs such as lymph node, spleen, and Peyer's patches. In the periphery, DCs are efficient antigen processing cells but are limited in their capacity to activate naive T cells. Upon activation (injury, inflammation, infection), DCs enter their final stage of maturation during which they downregulate the capacity to process new antigens, migrate out of the periphery into the secondary lymphoid organs, and acquire an extremely potent capacity to activate naive T cells. Several factors, such as cross-linking the CD40 surface molecules or the presence of TNF-α, can induce this final stage of maturation. CD40 is a type I integral membrane glycoprotein belonging to the TNF-receptor family. It is expressed on all mature B lymphocytes, dendritic cells, and some epithelial cells. Antibodies specific for CD40 molecules can induce proliferation of B cells when presented with interleukin-4 (IL-4) or antibodies specific for CD20 molecules. Also, stimulation of B cells with anti-CD40 antibodies and IL-4 can induce the switch of immunoglobulin production to the IgE isotype.

Steroid Hormones

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carrry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catcholamines in the central nervous system, and reduce inflammation. The principal

mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and autoimmune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Medroxyprogesterone (MAH), also known as 6α-methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

Mifepristone, also known as RU-486, is an antiprogesterone drug that blocks receptors of progesterone. It counteracts the effects of progesterone, which is needed to sustain pregnancy. Mifepristone induces spontaneous abortion when administered in early pregnancy followed by treatment with the prostaglandin misoprostol. Further studies show that mifepristone at a substantially lower dose can be highly effective as a postcoital contraceptive when administered within five days

after unprotected intercourse, thus providing women with a "morning-after pill" in case of contraceptive failure or sexual assault. Mifepristone also has potential uses in the treatment of breast and ovarian cancers in cases in which tumors are progesterone-dependent. It interferes with steroid-dependent growth of brain meningiomas, and may be useful in treatment of endometriosis where it blocks the estrogen-dependent growth of endometrial tissues. It may also be useful in treatment of uterine fibroid tumors and Cushing's Syndrome. Mifepristone binds to glucocorticoid receptors and interferes with cortisol binding. Mifepristone also may act as an anti-glucocorticoid and be effective for treating conditions where cortisol levels are elevated such as AIDS, anorexia nervosa, ulcers, diabetes, Parkinson's disease, multiple sclerosis, and Alzheimer's disease.

Danazol is a synthetic steroid derived from ethinyl testosterone. Danazol indirectly reduces estrogen production by lowering pituitary synthesis of follicle-stimulating hormone and LH. Danazol also binds to sex hormone receptors in target tissues, thereby exhibiting anabolic, antiestrognic, and weakly androgenic activity. Danazol does not possess any progestogenic activity, and does not suppress normal pituitary release of corticotropin or release of cortisol by the adrenal glands. Danazol is used in the treatment of endometriosis to relieve pain and inhibit endometrial cell growth. It is also used to treat fibrocystic breast disease and hereditary angioedema.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of

prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A_2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of β-adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, enzymes, referred to collectively as 'ENZM' and individually as 'ENZM-1,' 'ENZM-2,' 'ENZM-3,' 'ENZM-4,' 'ENZM-5,' 'ENZM-6,' 'ENZM-7,' 'ENZM-8,' 'ENZM-9,' 'ENZM-10,' 'ENZM-11,' 'ENZM-12,' 'ENZM-13,' 'ENZM-14,' 'ENZM-15,' 'ENZM-16,' 'ENZM-17,' 'ENZM-18,' 'ENZM-19,' 'ENZM-20,' 'ENZM-21,' 'ENZM-22,' 'ENZM-23,' 'ENZM-24,' 'ENZM-25,' 'ENZM-26,' 'ENZM-27,' 'ENZM-28,' 'ENZM-29,' 'ENZM-30,' 'ENZM-31,' 'ENZM-32,' 'ENZM-33,' 'ENZM-34,' 'ENZM-35,' 'ENZM-36,' 'ENZM-37,' and 'ENZM-38' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified enzymes and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified enzymes

and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-38.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-38. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:39-76.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least

90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to

said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active

fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the

activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is

for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"ENZM" refers to the amino acid sequences of substantially purified ENZM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of ENZM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

An "allelic variant" is an alternative form of the gene encoding ENZM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding ENZM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ENZM or a polypeptide with at least one functional characteristic of ENZM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

the polynucleotide encoding ENZM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding ENZM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ENZM. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of ENZM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of ENZM. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind ENZM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis

or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic ENZM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding ENZM or fragments of ENZM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an

exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of ENZM or a polynucleotide encoding ENZM which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:39-76 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:39-76, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:39-76 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:39-76 from related polynucleotides. The precise length of a fragment of SEQ ID NO:39-76 and the region of SEQ ID NO:39-76 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-38 is encoded by a fragment of SEQ ID NO:39-76. A fragment of SEQ ID NO:1-38 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-38. For example, a fragment of SEQ ID NO:1-38 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-38. The precise length of a fragment of SEQ ID NO:1-38 and the region of SEQ ID NO:1-38 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise

comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about

1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of ENZM which is

capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of ENZM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of ENZM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ENZM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an ENZM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of ENZM.

"Probe" refers to nucleic acids encoding ENZM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic

acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned

nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing ENZM, nucleic acids encoding ENZM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA,

in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell,

by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%,

or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human enzymes (ENZM), the polynucleotides encoding ENZM, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows

analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are enzymes. For example, SEQ ID NO:3 is 100% identical, from residue M1 to residue I539, to human acetolactate synthase (GenBank ID g2725625) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.0E-287, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also has homology to proteins that are localized to the cytoplasm and are lyases, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:3 also contains a thiamine pyrophosphate enzyme domain and a thiamine pyrophosphate enzyme N-terminus domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains, and a 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based TIGRFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:3 is an acetolactate synthase. In another example, SEQ ID NO:20 is 90% identical, from residue W142 to residue Y293, to human aldose reductase-like peptide (GenBank ID g3150035) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.9E-146, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:20 also has homology to human small intestine reductase, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:20 also contains an aldo/keto reductase family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:20 is an aldo/keto reductase. SEQ ID NO:1-2, SEQ ID NO:4-19, and SEQ ID NO:21-38 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-38 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide

SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:39-76 or that distinguish between SEQ ID NO:39-76 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses ENZM variants. Various embodiments of ENZM variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the ENZM amino acid sequence, and can contain at least one functional or structural characteristic of ENZM.

Various embodiments also encompass polynucleotides which encode ENZM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:39-76, which encodes ENZM. The polynucleotide sequences of SEQ ID NO:39-76, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding ENZM. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding ENZM. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:39-76 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:39-76. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of ENZM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding ENZM. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding ENZM, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding ENZM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding ENZM. For example, a polynucleotide comprising a sequence of SEQ ID NO:52 and a polynucleotide comprising a sequence of SEQ ID NO:53 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:55 and a polynucleotide comprising a sequence of SEQ ID NO:56 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:63 and a polynucleotide comprising a sequence of SEQ ID NO:64 are splice variants of each other; a polynucleotide comprising a sequence

of SEQ ID NO:66, a polynucleotide comprising a sequence of SEQ ID NO:67 and a polynucleotide comprising a sequence of SEQ ID NO:68 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:72 and a polynucleotide comprising a sequence of SEQ ID NO:73 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of ENZM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding ENZM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring ENZM, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode ENZM and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring ENZM under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding ENZM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ENZM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode ENZM and ENZM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding ENZM or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:39-76 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding ENZM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode ENZM may be cloned in recombinant DNA molecules that direct expression of ENZM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express ENZM.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter ENZM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ENZM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to

selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding ENZM may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, ENZM itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of ENZM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active ENZM, the polynucleotides encoding ENZM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding ENZM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding ENZM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding ENZM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control

signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding ENZM and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding ENZM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding ENZM. For example, routine cloning, subcloning, and propagation of polynucleotides encoding ENZM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen).

Ligation of polynucleotides encoding ENZM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of ENZM are needed, e.g. for the production of antibodies, vectors which direct high level expression of ENZM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of ENZM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of ENZM. Transcription of polynucleotides encoding ENZM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding ENZM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses ENZM in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of

DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of ENZM in cell lines is preferred. For example, polynucleotides encoding ENZM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk^- and apr^- cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding ENZM is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding ENZM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ENZM under the control of a single promoter. Expression of the marker gene in response to induction or selection

usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding ENZM and that express ENZM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of ENZM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on ENZM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ENZM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding ENZM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding ENZM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ENZM may be designed to contain signal sequences which direct

secretion of ENZM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding ENZM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric ENZM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ENZM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ENZM encoding sequence and the heterologous protein sequence, so that ENZM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled ENZM may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

ENZM, fragments of ENZM, or variants of ENZM may be used to screen for compounds

that specifically bind to ENZM. One or more test compounds may be screened for specific binding to ENZM. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to ENZM. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of ENZM can be used to screen for binding of test compounds, such as antibodies, to ENZM, a variant of ENZM, or a combination of ENZM and/or one or more variants ENZM. In an embodiment, a variant of ENZM can be used to screen for compounds that bind to a variant of ENZM, but not to ENZM having the exact sequence of a sequence of SEQ ID NO:1-38. ENZM variants used to perform such screening can have a range of about 50% to about 99% sequence identity to ENZM, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to ENZM can be closely related to the natural ligand of ENZM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor ENZM (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to ENZM can be closely related to the natural receptor to which ENZM binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for ENZM which is capable of propagating a signal, or a decoy receptor for ENZM which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to ENZM, fragments of ENZM, or variants of ENZM. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of ENZM. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of ENZM.

In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of ENZM.

In an embodiment, anticalins can be screened for specific binding to ENZM, fragments of ENZM, or variants of ENZM. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit ENZM involves producing appropriate cells which express ENZM, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing ENZM or cell membrane fractions which contain ENZM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ENZM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with ENZM, either in solution or affixed to a solid support, and detecting the binding of ENZM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter

its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

ENZM, fragments of ENZM, or variants of ENZM may be used to screen for compounds that modulate the activity of ENZM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ENZM activity, wherein ENZM is combined with at least one test compound, and the activity of ENZM in the presence of a test compound is compared with the activity of ENZM in the absence of the test compound. A change in the activity of ENZM in the presence of the test compound is indicative of a compound that modulates the activity of ENZM. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising ENZM under conditions suitable for ENZM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ENZM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding ENZM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding ENZM may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding ENZM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding ENZM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress ENZM, e.g., by secreting ENZM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of ENZM and enzymes. In addition, examples of tissues expressing ENZM can be found in Table 6 and can also be found in Example XI. Therefore, ENZM appears to play a role in autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased ENZM expression or activity, it is desirable to decrease the expression or activity of ENZM. In the treatment of disorders associated with decreased ENZM expression or activity, it is desirable to increase the expression or activity of ENZM.

Therefore, in one embodiment, ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an

infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis, rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple

sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell

proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ENZM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of ENZM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those listed above.

In a further embodiment, an antagonist of ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds ENZM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express ENZM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described

above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of ENZM may be produced using methods which are generally known in the art. In particular, purified ENZM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ENZM. Antibodies to ENZM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with ENZM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ENZM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of ENZM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to ENZM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the

splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ENZM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for ENZM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ENZM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ENZM epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ENZM. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of ENZM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple ENZM epitopes, represents the average affinity, or avidity, of the antibodies for ENZM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular ENZM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a

ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the ENZM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ENZM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of ENZM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding ENZM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ENZM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding ENZM (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding ENZM may be used for

somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in ENZM expression or regulation causes disease, the expression of ENZM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in ENZM are treated by constructing mammalian expression vectors encoding ENZM and introducing these vectors by mechanical means into ENZM-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of ENZM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). ENZM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc.

Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding ENZM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to ENZM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ENZM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding ENZM to cells which have one or more genetic abnormalities with respect to the expression of ENZM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding ENZM to target cells which have one or more genetic abnormalities with respect to the expression of ENZM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ENZM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding ENZM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,

resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for ENZM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ENZM-coding RNAs and the synthesis of high levels of ENZM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of ENZM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding ENZM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary

oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding ENZM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell.

Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a

compound which is effective in altering expression of a polynucleotide encoding ENZM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased ENZM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ENZM may be therapeutically useful, and in the treatment of disorders associated with decreased ENZM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ENZM may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ENZM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding ENZM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ENZM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic

acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ENZM, antibodies to ENZM, and mimetics, agonists, antagonists, or inhibitors of ENZM.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising ENZM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ENZM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ENZM or fragments thereof, antibodies of ENZM, and agonists, antagonists or inhibitors of ENZM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and

methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind ENZM may be used for the diagnosis of disorders characterized by expression of ENZM, or in assays to monitor patients being treated with ENZM or agonists, antagonists, or inhibitors of ENZM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ENZM include methods which utilize the antibody and a label to detect ENZM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring ENZM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ENZM expression. Normal or standard values for ENZM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ENZM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ENZM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding ENZM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of ENZM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ENZM, and to monitor regulation of ENZM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding ENZM or closely related molecules may be used to identify nucleic acid sequences which encode ENZM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding ENZM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ENZM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:39-76 or from genomic sequences including promoters, enhancers, and introns of the ENZM gene.

Means for producing specific hybridization probes for polynucleotides encoding ENZM include the cloning of polynucleotides encoding ENZM or ENZM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding ENZM may be used for the diagnosis of disorders associated with expression of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis,

rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders,

peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. Polynucleotides encoding ENZM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR

technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ENZM expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding ENZM may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding ENZM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding ENZM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of ENZM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ENZM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ENZM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding ENZM, or a fragment of a polynucleotide complementary to the polynucleotide encoding ENZM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding ENZM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding ENZM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the

anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of ENZM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, ENZM, fragments of ENZM, or antibodies specific for ENZM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by

hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are

indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for ENZM to quantify the levels of ENZM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by contacting the microarray with the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be

useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding ENZM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-

355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding ENZM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, ENZM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between ENZM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with ENZM, or fragments thereof, and washed. Bound ENZM is then detected by methods well known in the art. Purified ENZM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,

non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ENZM specifically compete with a test compound for binding ENZM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ENZM.

In additional embodiments, the nucleotide sequences which encode ENZM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/387,119, and U.S. Ser. No. 60/390,662, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA). Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP

vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using

reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite,

hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:39-76. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative enzymes are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode enzymes, the encoded polypeptides are analyzed by querying against PFAM models for enzymes. Potential enzymes are also identified by homology to Incyte cDNA sequences that have been annotated as enzymes. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the

sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases

using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of ENZM Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:39-76 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:39-76 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et

al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding ENZM are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled,

and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding ENZM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of ENZM Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in ENZM Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:39-76 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or

somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:39-76 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20

minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about

6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot

is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

SEQ ID NO:49 was downregulated in breast cancer tissue versus normal breast tissue as determined by microarray analysis. Expression of SEQ ID NO:49 was decreased in diseased tissue as compared with normal tissue from the same donor. Therefore, SEQ ID NO:49 can be used in monitoring treatment of, and diagnostic assays for, breast cancer.

SEQ ID NO:39 and SEQ ID NO:54 were downregulated in colon cancer tissue versus normal colon tissue as determined by microarray analysis. Expression of SEQ ID NO:39 and SEQ ID NO:54 was decreased in comparison of normal tissue from a donor with diseased tissue from the same donor. Therefore, SEQ ID NO:39 and SEQ ID NO:54 can be used in monitoring treatment of, and diagnostic assays for, colon cancer.

SEQ ID NO:48 and SEQ ID NO:51 were upregulated in prostate cancer cells versus normal prostate epithelial cells as determined by microarray analysis. Expression profiles of the prostate carcinoma lines LNCaP, PC-3, and DU 145 were compared to the expression profile of prostate epithelial cells (PrECs). DU 145 is a prostate carcinoma cell line isolated from a metastatic site in the brain of a donor with widespread metastatic prostate carcinoma. LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of donor with metastatic prostate carcinoma. PC-3 is a prostate adenocarcinoma cell line that was isolated from a metastatic site in the bone of a donor with grade IV prostate adenocarcinoma. Expression of SEQ ID NO:48 was increased in all three cell lines tested. Expression of SEQ ID NO:51 was increased in DU 145 and LNCaP cells. Therefore, SEQ ID NO:48 and SEQ ID NO:51 can be used in monitoring treatment of, and diagnostic assays for, prostate cancer.

SEQ ID NO:49, SEQ ID NO:54, SEQ ID NO:55, and SEQ ID NO:56 were differentially regulated in lung cancer tissue versus normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:49 and SEQ ID NO:54 was decreased in diseased tissue as compared with normal tissue from the same donor. Expression of SEQ ID NO:55 and SEQ ID NO:56 was increased in diseased tissue as compared with normal tissue from the same donor. Therefore, SEQ ID NO:49, SEQ ID NO:54, SEQ ID NO:55, and SEQ ID NO:56 can be used in monitoring treatment of, and

diagnostic assays for, lung cancer.

SEQ ID NO:49 was downregulated in ovarian cancer tissue versus normal ovarian tissue as determined by microarray analysis. Expression of SEQ ID NO:49 was decreased in comparisons of normal and diseased tissue from the same donor. Therefore, SEQ ID NO:49 can be used in monitoring treatment of, and diagnostic assays for, ovarian cancer.

SEQ ID NO:42 and SEQ ID NO:46 were differentially regulated in brain tissue associated with Alzheimer's disease (AD) versus normal brain tissue as determined by microarray analysis. In one experiment, specific dissected brain regions from the brain of a female with severe AD were compared to dissected regions from a normal female and two normal male brains. The diagnosis of normal or severe AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. Expression of SEQ ID NO:42 was decreased in tissue affected by severe AD as compared with normal brain tissue. In a second experiment, specific dissected brain regions from the brain of a female with mild AD were compared to from a normal donor were compared to specific dissected brain regions from two normal male and a normal female donor. The diagnosis of normal or mild AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. Expression of SEQ ID NO:46 was increased in tissue affected by mild AD compared with normal brain tissue. Therefore, SEQ ID NO:42 and SEQ ID NO:46 can be used in monitoring treatment of, and diagnostic assays for, Alzheimer's disease.

SEQ ID NO:55 and SEQ ID NO:56 were downregulated in tissue associated with Tangier disease versus normal tissue as determined by microarray analysis. Human fibroblasts were obtained from skin explants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable marker. In addition, both types of cells were cultured in the presence of cholesterol and compared with the same cell type cultured in the absence of cholesterol. TD derived cells are shown to be deficient in an assay of apoA-I mediated tritiated cholesterol efflux. Expression of SEQ ID NO:55 and SEQ ID NO:56 was decreased in fibroblasts from both patients with Tangier disease. Therefore, SEQ ID NO:55 and SEQ ID NO:56 can be used in monitoring treatment of, and diagnostic assays for, Tangier disease.

SEQ ID NO:45 were differentially regulated in C3A cells treated with gemfibrozil versus untreated C3A cells, as determined by microarray analysis. Early confluent C3A cells were treated with various amounts of Gemfibrozil (120, 600, 800, and 1200 μg/ml) dissolved in CMC for 1, 3, and 6 hours. Parallel samples of C3A cells were treated with 1% CMC only, as a control. Expression of

SEQ ID NO:45 was increased in C3A cells treated with gemfibrozil. Therefore, SEQ ID NO:45 can be used in monitoring treatment of, and diagnostic assays for, metabolic, cardiovascular, and liver disorders.

SEQ ID NO:55 and SEQ ID NO:56 were upregulated in C3A cells treated with a variety of steroids versus untreated C3A cells, as determined by microarray analysis. Early confluent C3A cells were treated for 1, 3, and 6 hours at concentrations of 1, 10, and 100 μM. Treated cells were compared to untreated early confluent C3A cells. Expression of SEQ ID NO:55 and SEQ ID NO:56 was increased in C3A cells treated with progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, and betamethasone. Therefore, SEQ ID NO:55 and SEQ ID NO:56 can be used for the diagnosis and monitoring of liver, endocrine, and reproductive diseases and in the diagnosis of and as a therapeutic target for inflammatory diseases and humoral immune response.

In another example, SEQ ID NO:60 showed differential expression in prostate cancer cells versus normal prostate epithelial cells, as determined by microarray analysis. Primary prostate epithelial cells were compared with prostate carcinomas representative of the different stages of tumor progression. Cell lines compared included: a)PrEC, a primary prostate epithelial cell line isolated from a normal donor, b) DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of a donor with widespread metastatic prostate carcinoma, c) LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a donor with metastatic prostate carcinoma, and d) PC-3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a donor with grade IV prostate adenocarcinoma. Cells grown under restrictive conditions were compared to normal PrECs grown under restrictive conditions. Expression of SEQ ID NO:60 was increased in all LNCaP samples and in 3 of 4 DU 145 samples tested. Therefore, SEQ ID NO:60 can be used in monitoring treatment of, and diagnostic assays for, prostate cancer.

In a further example, SEQ ID NO:75 showed differential expression in fibroblasts with Tangier disease versus normal fibroblasts, as determined by microarray analysis. Human fibroblasts were obtained from skin explants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable marker. In addition, both types of cells were cultured in the presence of cholesterol and compared with the same cell type cultured in the absence of cholesterol. TD-derived cells are deficient in an assay of apoA-I mediated tritiated cholesterol efflux. Expression of SEQ ID NO:75 was increased by an average of more than twofold in fibroblasts obtained from donors with Tangier disease. Therefore, SEQ ID NO:75 can be used in diagnostic

assays for Tangier disease.

In addition, SEQ ID NO:39, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:49, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:61, and SEQ ID NO:69 showed tissue-specific expression. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another.

The expression of SEQ ID NO:39 was increased by at least two-fold in occipital cortex and in temporal cortex brain tissue, as compared to the reference sample. Therefore, SEQ ID NO:39 can be used as a tissue marker for brain.

The expression of SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:49, SEQ ID NO:54, SEQ ID NO:55, and SEQ ID NO:56 was increased by at least two-fold in liver as compared to the reference sample. Therefore, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:49, SEQ ID NO:54, SEQ ID NO:55, and SEQ ID NO:56 can be used as a tissue marker for liver.

The expression of SEQ ID NO:54 and SEQ ID NO:45 was increased by at least two-fold in kidney as compared to the reference sample. Therefore, SEQ ID NO:54 and SEQ ID NO:45 can be used as a tissue marker for kidney.

The expression of SEQ ID NO:54 was increased by at least two-fold in duodenum as compared to the reference sample. Therefore, SEQ ID NO:54 can be used as a tissue marker for duodenum.

The expression of SEQ ID NO:55 and SEQ ID NO:56 was increased by at least two-fold in psoas skeletal muscle as compared to the reference sample. Therefore, SEQ ID NO:55 and SEQ ID NO:56 can be used as a tissue marker for psoas skeletal muscle.

The expression of SEQ ID NO:42 was increased by at least two-fold in brain as compared to the reference sample. Therefore, SEQ ID NO:42 can be used as a tissue marker for brain.

The expression of SEQ ID NO:55 and SEQ ID NO:56 was increased by at least two-fold in left ventricle and right ventricle as compared to the reference sample. Therefore, SEQ ID NO:55 and SEQ ID NO:56 can be used as a tissue marker for left ventricle and right ventricle.

The expression of SEQ ID NO:61 was increased by at least two-fold in right ventricle tissue

and adrenal tissue, as compared to the reference sample. Therefore, SEQ ID NO:61 can be used as a tissue marker for right ventricle tissue and adrenal tissue.

The expression of SEQ ID NO:69 was increased by at least two-fold in liver and kidney tissue, as compared to the reference sample. Therefore, SEQ ID NO:69 can be used as a tissue marker for liver and kidney tissue.

XII. Complementary Polynucleotides

Sequences complementary to the ENZM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ENZM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of ENZM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the ENZM-encoding transcript.

XIII. Expression of ENZM

Expression and purification of ENZM is achieved using bacterial or virus-based expression systems. For expression of ENZM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express ENZM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ENZM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ENZM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, ENZM is synthesized as a fusion protein with, e.g., glutathione S-

transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from ENZM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified ENZM obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

ENZM function is assessed by expressing the sequences encoding ENZM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of ENZM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ENZM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding ENZM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of ENZM Specific Antibodies

ENZM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the ENZM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-ENZM activity by, for example, binding the peptide or ENZM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring ENZM Using Specific Antibodies

Naturally occurring or recombinant ENZM is substantially purified by immunoaffinity chromatography using antibodies specific for ENZM. An immunoaffinity column is constructed by covalently coupling anti-ENZM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing ENZM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ENZM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

antibody/ENZM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ENZM is collected.

XVII. Identification of Molecules Which Interact with ENZM

ENZM, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ENZM, washed, and any wells with labeled ENZM complex are assayed. Data obtained using different concentrations of ENZM are used to calculate values for the number, affinity, and association of ENZM with the candidate molecules.

Alternatively, molecules interacting with ENZM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

ENZM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of ENZM Activity

ENZM activity is demonstrated through a variety of specific enzyme assays; some of which are outlined below.

ENZM oxidoreductase activity is measured by the increase in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of oxidation activity, or the decrease in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of reduction activity (Dalziel, K. (1963) J. Biol. Chem. 238:2850-2858). One of three substrates may be used: Asn-βGal, biocytidine, or ubiquinone-10. The respective subunits of the enzyme reaction, for example, cytochrome c_1-b oxidoreductase and cytochrome c, are reconstituted. The reaction mixture contains a)1-2 mg/ml ENZM; and b) 15 mM substrate, 2.4 mM $NAD(P)^+$ in 0.1 M phosphate buffer, pH 7.1 (oxidation reaction), or 2.0 mM NAD(P)H, in 0.1 M Na_2HPO_4 buffer, pH 7.4 (reduction reaction); in a total volume of 0.1 ml. Changes in absorbance at 340 nm (A_{340}) are measured at 23.5°C using a recording spectrophotometer (Shimadzu Scientific Instruments, Inc., Pleasanton, CA). The amount of NAD(P)H is stoichiometrically equivalent to the amount of substrate initially present, and the change in A_{340} is a direct measure of the amount of NAD(P)H produced; $\Delta A_{340} = 6620[NADH]$. ENZM activity is proportional to the amount of NAD(P)H present in the assay.

Aldo/keto reductase activity of ENZM is proportional to the decrease in absorbance at 340 nm as NADPH is consumed (or increased absorbance if NADPH is produced, i.e., if the reverse reaction is monitored). A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 mg ENZM and an appropriate level of substrate. The reaction is incubated at 30°C and the reaction is monitored continuously with a spectrophotometer. ENZM activity is calculated as mol NADPH consumed / mg of ENZM.

Acyl-CoA dehydrogenase activity of ENZM is measured using an anaerobic electron transferring flavoprotein (ETF) assay. The reaction mixture comprises 50 mM Tris-HCl (pH 8.0), 0.5% glucose, and 50 μM acyl-CoA substrate (i.e., isovaleryl-CoA) that is pre-warmed to 32 °C. The mixture is depleted of oxygen by repeated exposure to vacuum followed by layering with argon. Trace amounts of oxygen are removed by the addition of glucose oxidase and catalase followed by the addition of ETF to a final concentration of 1 μM. The reaction is initiated by addition of purified ENZM or a sample containing ENZM and exciting the reaction at 342 nm. Quenching of fluorescence caused by the transfer of electrons from the substrate to ETF is monitored at 496 nm. 1 unit of acyl-CoA dehydrogenase activity is defined as the amount of ENZM required to reduce 1 μmol of ETF per minute (Reinard, T. et al. (2000) J. Biol. Chem. 275:33738-33743).

Alcohol dehydrogenase activity of ENZM is measured by following the conversion of NAD^+ to NADH at 340 nm ($\varepsilon_{340} = 6.22\ mM^{-1}\ cm^{-1}$) at 25°C in 0.1 M potassium phosphate (pH 7.5), 0.1 M glycine (pH 10.0), and 2.4 mM NAD^+. Substrate (e.g., ethanol) and ENZM are then added to the reaction. The production of NADH results in an increase in absorbance at 340 nm and correlates with the oxidation of the alcohol substrate and the amount of alcohol dehydrogenase activity in the ENZM sample (Svensson, S. (1999) J. Biol. Chem. 274:29712-29719).

Aldehyde dehydrogenase activity of ENZM is measured by determining the total hydrolase + dehydrogenase activity of ENZM and subtracting the hydrolase activity. Hydrolase activity is first determined in a reaction mixture containing 0.05 M Tris-HCl (pH 7.8), 100 mM 2-mercaptoethanol, and 0.5-18 μM substrate, e.g., 10-HCO-HPteGlu (10-formyltetrahydrofolate; HPteGlu, tetrahydrofolate) or 10-FDDF (10-formyl-5,8-dideazafolate). Approximately 1 μg of ENZM is added in a final volume of 1.0 ml. The reaction is monitored and read against a blank cuvette, containing all components except enzyme. The appearance of product is measured at either 295 nm for 5,8-dideazafolate or 300 nm for HPteGlu using molar extinction coefficients of 1.89×10^4 and 2.17×10^4 for 5,8-dideazafolate and HPteGlu, respectively. The addition of $NADP^+$ to the reaction mixture allows the measurement of both dehydrogenase and hydrolase activity (assays are performed as before). Based on the production of product in the presence of $NADP^+$ and the production of product

in the absence of the cofactor, aldehyde dehydrogenase activity is calculated for ENZM. In the alternative, aldehyde dehydrogenase activity is assayed using propanal as substrate. The reaction mixture contains 60 mM sodium pyrophosphate buffer (pH 8.5), 5 mM propanal, 1 mM $NADP^+$, and ENZM in a total volume of 1 ml. Activity is determined by the increase in absorbance at 340 nm, resulting from the generation of NADPH, and is proportional to the aldehyde dehydrogenase activity in the sample (Krupenko, S.A. et al. (1995) J. Biol. Chem. 270:519-522).

6-phosphogluconate dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in 120 mM triethanolamine (pH 7.5), 0.1 mM EDTA, 0.5 mM $NADP^+$, and 10-150 μM 6-phosphogluconate as substrate at 20-25°C. The production of NADPH is measured fluorimetrically (340 nm excitation, 450 nm emission) and is indicative of 6-phosphogluconate dehydrogenase activity. Alternatively, the production of NADPH is measured photometrically, based on absorbance at 340 nm. The molar amount of NADPH produced in the reaction is proportional to the 6-phosphogluconate dehydrogenase activity in the sample (Tetaud et al., *supra*).

Ribonucleotide diphosphate reductase activity of ENZM is determined by incubating purified ENZM, or a composition comprising ENZM, along with dithiothreitol, Mg^{++}, and ADP, GDP, CDP, or UDP substrate. The product of the reaction, the corresponding deoxyribonucleotide, is separated from the substrate by thin-layer chromatography. The reaction products can be distinguished from the reactants based on rates of migration. The use of radiolabeled substrates is an alternative for increasing the sensitivity of the assay. The amount of deoxyribonucleotides produced in the reaction is proportional to the amount of ribonucleotide diphosphate reductase activity in the sample (note that this is true only for pre-steady state kinetic analysis of ribonucleotide diphosphate reductase activity, as the enzyme is subject to negative feedback inhibition by products) (Nutter and Cheng, *supra*).

Dihydrodiol dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in a reaction mixture comprising 50 mM glycine (pH 9.0), 2.3 mM $NADP^+$, 8% DMSO, and a trans-dihydrodiol substrate, selected from the group including but not limited to, (±)-trans-naphthalene-1,2-dihydrodiol, (±)-trans-phenanthrene-1,2-dihydrodiol, and (±)-trans-chrysene-1,2-dihydrodiol. The oxidation reaction is monitored at 340 nm to detect the formation of NADPH, which is indicative of the oxidation of the substrate. The reaction mixture can also be analyzed before and after the addition of ENZM by circular dichroism to determine the stereochemistry of the reaction components and determine which enantiomers of a racemic substrate composition are oxidized by the ENZM (Penning, *supra*).

Glutathione S-transferase (GST) activity of ENZM is determined by measuring the ENZM

catalyzed conjugation of GSH with 1-chloro-2,4-dinitrobenzene (CDNB), a common substrate for most GSTs. ENZM is incubated with 1 mM CDNB and 2.5 mM GSH together in 0.1M potassium phosphate buffer, pH 6.5, at 25°C. The conjugation reaction is measured by the change in absorbance at 340 nm using an ultraviolet spectrophotometer. ENZM activity is proportional to the change in absorbance at 340 nm.

15-oxoprostaglandin 13-reductase (PGR) activity of ENZM is measured following the separation of contaminating 15-hydroxyprostaglandin dehydrogenase (15-PGDH) activity by DEAE chromatography. Following isolation of PGR containing fractions (or using the purified ENZM), activity is assayed in a reaction comprising 0.1 M sodium phosphate (pH 7.4), 1 mM 2-mercaptoethanol, 20 μg substrate (e.g., 15-oxo derivatives of prostaglandins PGE_1, PGE_2, and $PGE_{2\alpha}$), and 1 mM NADH (or a higher concentration of NADPH). ENZM is added to the reaction which is then incubated for 10 min at 37°C before termination by the addition of 0.25 ml 2 N NaOH. The amount of 15-oxo compound remaining in the sample is determined by measuring the maximum absorption at 500 nm of the terminated reaction and comparing this value to that of a terminated control reaction that received no ENZM. 1 unit of enzyme is defined as the amount required to catalyze the oxidation of 1 μmol substrate per minute and is proportional to the amount of PGR activity in the sample.

Choline dehydrogenase activity of ENZM is identified by the ability of *E. coli*, transformed with an ENZM expression vector, to grow on media containing choline as the sole carbon and nitrogen source. The ability of the transformed bacteria to thrive is indicative of choline dehydrogenase activity (Magne Østerås, M. (1998) Proc. Natl. Acad. Sci. USA 95:11394-11399).

ENZM thioredoxin activity is assayed as described (Luthman, M. (1982) Biochemistry 21:6628-6633). Thioredoxins catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges. One way to measure the thiol:disulfide exchange is by measuring the reduction of insulin in a mixture containing 0.1 M potassium phosphate, pH 7.0, 2 mM EDTA, 0.16 μM insulin, 0.33 mM DTT, and 0.48 mM NADPH. Different concentrations of ENZM are added to the mixture, and the reaction rate is followed by monitoring the oxidation of NADPH at 340 nM.

ENZM transferase activity is measured through assays such as a methyl transferase assay in which the transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate is measured (Bokar, J.A. et al. (1994) J. Biol. Chem. 269:17697-17704). Reaction mixtures (50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5 μCi [*methyl*-^{3}H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg ENZM, and

acceptor substrate (0.4 μg [^{35}S]RNA or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then at 65°C for 5 minutes. The products are separated by chromatography or electrophoresis and the level of methyl transferase activity is determined by quantification of *methyl*-^{3}H recovery.

Aminotransferase activity of ENZM is assayed by incubating samples containing ENZM for 1 hour at 37°C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200 μl of 150 mM Tris acetate buffer (pH 8.0) containing 70 μM PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative, L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid and xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli et al., *supra*).

In another alternative, aminotransferase activity of ENZM is measured by determining the activity of purified ENZM or crude samples containing ENZM toward various amino and oxo acid substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, pyridoxal 5'-phosphate (PLP). The reactions are performed at 25°C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 μM purified ENZM or ENZM containing samples and substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to pyridoxamine 5' phosphate (PMP). The specificity and relative activity of ENZM is determined by the activity of the enzyme preparation against specific substrates (Vacca, *supra*).

ENZM chitinase activity is determined with the fluorogenic substrates 4-methylumbelliferyl chitotriose, methylumbelliferyl chitobiose, or methylumbelliferyl N-acetylglucosamine. Purified ENZM is incubated with 0.5uM substrate at pH 4.0 (0.1M citrate buffer), pH 5.0 (0.1M phosphate buffer), or pH 6.0 (0.1M Tris-HCL). After various times of incubation, the reaction is stopped by the addition of 0.1M glycine buffer, pH 10.4, and the concentration of free methylumbelliferone is determined fluorometrically. Chitinase B from *Serratia marcescens* may be used as a positive control (Hakala, *supra*).

ENZM isomerase activity is determined by measuring 2-hydroxyhepta-2,4-diene,1,7 dioate isomerase (HHDD isomerase) activity, as described by Garrido-Peritierra, A. and R.A. Cooper (1981; Eur. J. Biochem. 17:581-584). The sample is combined with 5-carboxymethyl-2-oxo-hex-3-ene-1,5, dioate (CMHD), which is the substrate for HHDD isomerase. CMHD concentration is monitored by

measuring its absorbance at 246 nm. Decrease in absorbance at 246 nm is proportional to HHDD isomerase activity of ENZM.

ENZM isomerase activity such as peptidyl prolyl *cis/trans* isomerase activity can be assayed by an enzyme assay described by Rahfeld (*supra*). The assay is performed at 10°C in 35 mM HEPES buffer, pH 7.8, containing chymotrypsin (0.5 mg/ml) and ENZM at a variety of concentrations. Under these assay conditions, the substrate, Suc-Ala-Xaa-Pro-Phe-4-NA, is in equilibrium with respect to the prolyl bond, with 80-95% in *trans* and 5-20% in *cis* conformation. An aliquot (2 μl) of the substrate dissolved in dimethyl sulfoxide (10 mg/ml) is added to the reaction mixture described above. Only the *cis* isomer is a substrate for cleavage by chymotrypsin. Thus, as the substrate is isomerized by ENZM, the product is cleaved by chymotrypsin to produce 4-nitroanilide, which is detected by its absorbance at 390 nm. 4-Nitroanilide appears in a time-dependent and a ENZM concentration-dependent manner.

Alternatively, peptidyl prolyl *cis-trans* isomerase activity of ENZM can be assayed using a chromogenic peptide in a coupled assay with chymotrypsin (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111).

UDP glucuronyltransferase activity of ENZM is measured using a colorimetric determination of free amine groups (Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London). An amine-containing substrate, such as 2-aminophenol, is incubated at 37°C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM $MgCl_2$, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylene diamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm, for example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Adenylosuccinate synthetase activity of ENZM is measured by synthesis of AMP from IMP. The sample is combined with AMP. IMP concentration is monitored spectrophotometrically at 248 nm at 23°C (Wang, W. et al. (1995) J. Biol. Chem. 270:13160-13163). The increase in IMP concentration is proportional to ENZM activity.

Alternatively, AMP binding activity of ENZM is measured by combining the sample with ^{32}P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to ENZM activity.

In another alternative, xenobiotic carboxylic acid:CoA ligase activity of ENZM is measured by combining the sample with γ-^{33}P-ATP and measuring the formation of γ-^{33}P- pyrophosphate with time (Vessey, D.A. et al. (1998) J. Biochem. Mol. Toxicol. 12:151-155).

Protein phosphatase (PP) activity can be measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). ENZM is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62).

Alternatively, acid phosphatase activity of ENZM is demonstrated by incubating ENZM containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of ENZM in the assay.

In the alternative, ENZM activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM ENZM in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% 2-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × g for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

The adenosine deaminase activity of ENZM is determined by measuring the rate of deamination that occurs when adenosine substrate is incubated with ENZM. Reactions are performed with a predetermined amount of ENZM in a final volume of 3.0 ml containing 53.3 mM potassium phosphate and 0.045 mM adenosine. Assay reagents excluding ENZM are mixed in a quartz cuvette and equilibrated to 25° C. Reactions are initiated by the addition of ENZM and are mixed immediately by inversion. The decrease in light absorbance at 265 nm resulting from the hydrolysis of adenosine to inosine is measured using a spectrophotometer. The decrease in the $A_{265\ nm}$ is recorded for approximately 5 minutes. The decrease in light absorbance is proportional to the activity of ENZM in

the assay.

ENZM hydrolase activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon and Bond, *supra*, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase).

An assay for carbonic anhydrase activity of ENZM uses the fluorescent pH indicator 8-hydroxypyrene-1,3,6-trisulfonate (pyranine) in combination with stopped-flow fluorometry to measure carbonic anhydrase activity (Shingles, et al. 1997, Anal. Biochem. 252:190-197). A pH 6.0 solution is mixed with a pH 8.0 solution and the initial rate of bicarbonate dehydration is measured. Addition of carbonic anhydrase to the pH 6.0 solution enables the measurement of the initial rate of activity at physiological temperatures with resolution times of 2 ms. Shingles et al. (*supra*) used this assay to resolve differences in activity and sensitivity to sulfonamides by comparing mammalian carbonic anhydrase isoforms. The fluorescent technique's sensitivity allows the determination of initial rates with a protein concentration as little as 65 ng/ml.

Decarboxylase activity of ENZM is measured as the release of CO_2 from labeled substrate. For example, ornithine decarboxylase activity of ENZM is assayed by measuring the release of CO_2 from L-[1-^{14}C]-ornithine (Reddy, S.G et al. (1996) J. Biol. Chem. 271:24945-24953). Activity is measured in 200 μl assay buffer (50 mM Tris/HCl, pH 7.5, 0.1 mM EDTA, 2 mM dithiothreitol, 5 mM NaF, 0.1% Brij35, 1 mM PMSF, 60 μM pyridoxal-5-phosphate) containing 0.5 mM L-ornithine plus 0.5 μCi L-[1-^{14}C]ornithine. The reactions are stopped after 15-30 minutes by addition of 1 M citric acid, and the $^{14}CO_2$ evolved is trapped on a paper disk filter saturated with 20 μl of 2 N NaOH. The radioactivity on the disks is determined by liquid scintillation spectography. The amount of $^{14}CO_2$ released is proportional to ornithine decarboxylase activity of ENZM.

AdoHCYase activity of ENZM in the hydrolytic direction is performed spectroscopically by measuring the rate of the product (homocysteine) formed by reaction with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). To 800 μl of an enzyme solution containing 4.7 μg of ENZM and 4 units of adenosine deaminase in 50 mM potassium phosphate buffer, pH 7.2, containing 1 mM EDTA (buffer A), is added 200 μl of *S*-Adenosyl-L-homocysteine (500 μM) containing 250 μM DTNB in buffer A. The reaction mixture is incubated at 37°C for 2 minutes. Hydrolytic activity is monitored at 412 nm continuously using a diode array UV spectrophotometer. Enzyme activity is defined as the amount of enzyme that can hydrolyze 1 μmol of *S*-Adenosyl-L-homocysteine/minute (Yuan, C-S et al. (1996) J.

Biol. Chem. 271:28009-28015).

AdoHCYase activity of ENZM can be measured in the synthetic direction as the production of S-adenosyl homocysteine using 3-deazaadenosine as a substrate (Sganga et al. *supra*). Briefly, ENZM is incubated in a 100 μl volume containing 0.1 mM 3-deazaadenosine, 5 mM homocysteine, 20 mM HEPES (pH 7.2). The assay mixture is incubated at 37°C for 15 minutes. The reaction is terminated by the addition of 10 μl of 3 M perchloric acid. After incubation on ice for 15 minutes, the mixture is centrifuged for 5 minutes at 18,000 x *g* in a microcentrifuge at 4°C. The supernatant is removed, neutralized by the addition of 1 M potassium carbonate, and centrifuged again. A 50 μl aliquot of supernatant is then chromatographed on an Altex Ultrasphere ODS column (5 μm particles, 4.6 x 250 mm) by isocratic elution with 0.2 M ammonium dihydrogen phosphate (Aldrich) at a flow rate of 1 ml/min. Protein is determined by the bicinchoninic acid assay (Pierce).

Alternatively, AdoHCYase activity of ENZM can be measured in the synthetic direction by a TLC method (Hershfield, M.S. et al. (1979) J. Biol. Chem. 254:22-25). In a preincubation step, 50 μM [8-^{14}C]adenosine is incubated with 5 molar equivalents of NAD$^+$ for 15 minutes at 22°C. Assay samples containing ENZM in a 50 μl final volume of 50 mM potassium phosphate buffer, pH 7.4, 1 mM DTT, and 5 mM homocysteine, are mixed with the preincubated [8-^{14}C]adenosine/NAD$^+$ to initiate the reaction. The reaction is incubated at 37°C, and 1 μl samples are spotted on TLC plates at 5 minute intervals for 30 minutes. The chromatograms are developed in butanol-1/glacial acetic acid/water (12:3:5, v/v) and dried. Standards are used to identify substrate and products under ultraviolet light. The complete spots containing [^{14}C]adenosine and [^{14}C]SAH are then detected by exposing x-ray film to the TLC plate. The radiolabeled substrate and product are then cut from the chromatograms and counted by liquid scintillation spectrometry. Specific activity of the enzyme is determined from the linear least squares slopes of the product *vs* time plots and the milligrams of protein in the sample (Bethin, K.E. et al. (1995) J. Biol. Chem. 270:20698-20702).

Asparaginase activity of ENZM can be measured in the hydrolytic direction by determining the amount of radiolabeled L-aspartate released from 0.6 mM *N*^{4}-β'-*N*-acetylglucosaminyl-L-asparagine substrate when it is incubated at 25°C with ENZM in 50 mM phosphate buffer, pH 7.5 (Kaartinen, V. et al. (1991) J. Biol. Chem. 266:5860-5869).

Acyl CoA Acid Hydrolase activity of ENZM in the hydrolytic direction is performed spectroscopically by monitoring the appearance of the product (CoASH) formed by reaction of substrate (acyl-CoA) and ENZM with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). The final reaction volume is 1 ml of 0.05 M potassium phosphate buffer, pH 8, containing 0.1 mM DTNB, 20 μg/ml bovine serum albumin, 10 μM of acyl-CoA of different lengths (C6-CoA, C10-CoA, C14-CoA and

C18-CoA, Sigma), and ENZM. The reaction mixture is incubated at 22°C for 7 minutes. Hydrolytic activity is monitored spectrophotometrically by measuring absorbance at 412 nm (Poupon, V. et al. (1999) J. Biol. Chem. 274:19188-19194).

ENZM activity of ENZM can be measured spectrophotometrically by determining the amount of solubilized RNA that is produced as a result of incubation of RNA substrate with ENZM. 5 μl (20 μg) of a 4 mg/ml solution of yeast tRNA (Sigma) is added to 0.8 ml of 40 mM sodium phosphate, pH 7.5, containing ENZM. The reaction is incubated at 25°C for 15 minutes. The reaction is stopped by addition of 0.5 ml of an ice-cold fresh solution of 20 mM lanthanum nitrate plus 3% perchloric acid. The stopped reaction is incubated on ice for at least 15 min, and the insoluble tRNA is removed by centrifugation for 5 min at 10,000 g. Solubilized tRNA is determined as UV absorbance (260 nm) of the remaining supernatant, with A_{260} of 1.0 corresponding to 40 μg of solubilized RNA (Rosenberg, H.F. et al. (1996) Nucleic Acids Research 24:3507-3513).

ENZM activity can be determined as the ability of ENZM to cleave ^{32}P internally labeled *T. thermophila* pre-tRNA^{Gln}. ENZM and substrate are added to reaction vessels and reactions are carried out in MBB buffer (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$) for 1 hour at 37°C. Reactions are terminated with the addition of an equal volume of sample loading buffer (SLB: 40 mM EDTA, 8 M urea, 0.2% xylene cyanol, and 0.2% bromophenol blue). The reaction products are separated by electrophoresis on 8 M urea, 6% polyacrylamide gels and analyzed using detection instruments and software capable of quantification of the products. One unit of ENZM activity is defined as the amount of enzyme required to cleave 10% of 28 fmol of *T. thermophila* pre-tRNA^{Gln} to mature products in 1 hour at 37°C (True, H.L. et al. (1996) J. Biol. Chem. 271:16559-16566).

Alternatively, cleavage of ^{32}P internally labeled substrate tRNA by ENZM can be determined in a 20 μl reaction mixture containing 30 mM HEPES-KOH (pH 7.6), 6 mM $MgCl_2$, 30 mM KCl, 2 mM DTT, 25 μg/ml bovine serum albumin, 1 unit/μl rRNasin, and 5,000-50,000 cpm of gel-purified substrate RNA. 3.0 μl of ENZM is added to the reaction mixture, which is then incubated at 37°C for 30 minutes. The reaction is stopped by guanidinium/phenol extraction, precipitated with ethanol in the presence of glycogen, and subjected to denaturing polyacrylamide gel electrophoresis (6 or 8% polyacrylamide, 7 M urea) and autoradiography (Rossmanith, W. et al. (1995) J. Biol. Chem. 270:12885-12891). The ENZM activity is proportional to the amount of cleavage products detected.

ENZM activity can be measured by determining the amount of free adenosine produced by the hydrolysis of AMP, as described by Sala-Newby et al., *supra*. Briefly, ENZM is incubated with AMP in a suitable buffer for 10 minutes at 37°C. Free adenosine is separated from AMP and measured by reverse phase HPLC.

Alternatively, ENZM activity is measured by the hydrolysis of ADP-ribosylarginine (Konczalik, P. and J. Moss (1999) J. Biol. Chem. 274:16736-16740). 50 ng of ENZM is incubated with 100 μM ADP-ribosyl-[^{14}C]arginine (78,000 cpm) in 50 mM potassium phosphate, pH 7.5, 5 mM dithiothreitol, 10 mM $MgCl_2$ in a final volume of 100 μl. After 1 h at 37° C, 90 μl of the sample is applied to a column (0.5 × 4 cm) of Affi-Gel 601 (boronate) equilibrated and eluted with five 1-ml portions of 0.1 M glycine, pH 9.0, 0.1 M NaCl, and 10 mM $MgCl_2$. Free ^{14}C-Arg in the total eluate is measured by liquid scintillation counting.

Epoxide hydrolase activity of ENZM can be determined with a radiometric assay utilizing [H^3]-labeled *trans*-stilbene oxide (TSO) as substrate. Briefly, ENZM is preincubated in Tris-HCl pH 7.4 buffer in a total volume of 100 μl for 1 minute at 37°C. 1 μl of [H^3]-labeled TSO (0.5 μM in EtOH) is added and the reaction mixture is incubated at 37°C for 10 minutes. The reaction mixture is extracted with 200 μl n-dodecane. 50 μl of the aqueous phase is removed for quantification of diol product in a liquid scintillation counter (LSC). ENZM activity is calculated as nmol diol product/min/mg protein (Gill, S.S. et al. (1983) Analytical Biochemistry 131:273-282).

Lysophosphatidic acid acyltransferase activity of ENZM is measured by incubating samples containing ENZM with 1 mM of the thiol reagent 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), 50 μm LPA, and 50 μm acyl-CoA in 100 mM Tris-HCl, pH 7.4. The reaction is initiated by addition of acyl-CoA, and allowed to reach equilibrium. Transfer of the acyl group from acyl-CoA to LPA releases free CoA, which reacts with DTNB. The product of the reaction between DTNB and free CoA absorbs at 413 nm. The change in absorbance at 413 nm is measured using a spectrophotometer, and is proportional to the lysophosphatidic acid acyltransferase activity of ENZM in the sample.

N-acyltransferase activity of ENZM is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. ENZM is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R. et al. (1990; J. Biol. Chem. 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholyl-CoA and ^{3}H-glycine or ^{3}H-taurine, separating the tritiated cholate conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA (CoASH) described below.

N-acetyltransferase activity of ENZM is measured using the transfer of radiolabel from [^{14}C]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) J. Neurochem.

24:1083-5). Alternatively, a newer spectrophotometric assay based on DTNB reaction with CoASH may be used. Free thiol-containing CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate. CoASH is detected using the absorbance of DTNB conjugate at 412 nm (De Angelis, J. et al. (1997) J. Biol. Chem. 273:3045-3050). ENZM activity is proportional to the rate of radioactivity incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Galactosyltransferase activity of ENZM is determined by measuring the transfer of galactose from UDP-galactose to a GlcNAc-terminated oligosaccharide chain in a radioactive assay. (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:58-65.) The ENZM sample is incubated with 14 μl of assay stock solution (180 mM sodium cacodylate, pH 6.5, 1 mg/ml bovine serum albumin, 0.26 mM UDP-galactose, 2 μl of UDP-[^{3}H]galactose), 1 μl of $MnCl_2$ (500 mM), and 2.5 μl of GlcNAcβO-$(CH_2)_8$-CO_2Me (37 mg/ml in dimethyl sulfoxide) for 60 minutes at 37°C. The reaction is quenched by the addition of 1 ml of water and loaded on a C18 Sep-Pak cartridge (Waters), and the column is washed twice with 5 ml of water to remove unreacted UDP-[^{3}H]galactose. The [^{3}H]galactosylated GlcNAcβO-$(CH_2)_8$-CO_2Me remains bound to the column during the water washes and is eluted with 5 ml of methanol. Radioactivity in the eluted material is measured by liquid scintillation counting and is proportional to galactosyltransferase activity of ENZM in the starting sample.

Phosphoribosyltransferase activity of ENZM is measured as the transfer of a phosphoribosyl group from phosphoribosylpyrophosphate (PRPP) to a purine or pyrimidine base. Assay mixture (20 µl) containing 50 mM Tris acetate, pH 9.0, 20 mM 2-mercaptoethanol, 12.5 mM $MgCl_2$, and 0.1 mM labeled substrate, for example, [^{14}C]uracil, is mixed with 20 µl of ENZM diluted in 0.1 M Tris acetate, pH 9.7, and 1 mg/ml bovine serum albumin. Reactions are preheated for 1 min at 37°C, initiated with 10 µl of 6 mM PRPP, and incubated for 5 min at 37°C. The reaction is stopped by heating at 100°C for 1 min. The product [^{14}C]UMP is separated from [^{14}C]uracil on DEAE-cellulose paper (Turner, R.J. et al. (1998) J. Biol. Chem. 273:5932-5938). The amount of [^{14}C]UMP produced is proportional to the phosphoribosyltransferase activity of ENZM.

ADP-ribosyltransferase activity of ENZM is measured as the transfer of radiolabel from adenine-NAD to agmatine (Weng, B. et al. (1999) J. Biol. Chem. 274:31797-31803). Purified ENZM is incubated at 30°C for 1 hr in a total volume of 300 µl containing 50 mM potassium phosphate (pH. 7.5), 20 mM agmatine, and 0.1 mM [adenine-U-^{14}C]NAD (0.05 mCi). Samples (100 µl) are applied to Dowex columns and [^{14}C]ADP-ribosylagmatine eluted with 5 ml of water for liquid scintillation counting. The amount of radioactivity recovered is proportional to ADP-ribosyltransferase activity of ENZM.

An ENZM activity assay measures aminoacylation of tRNA in the presence of a radiolabeled substrate. SYNT is incubated with [^{14}C]-labeled amino acid and the appropriate cognate tRNA (for example, [^{14}C]alanine and tRNA^{ala}) in a buffered solution. ^{14}C-labeled product is separated from free [^{14}C]amino acid by chromatography, and the incorporated ^{14}C is quantified using a scintillation counter. The amount of ^{14}C-labeled product detected is proportional to the activity of ENZM in this assay (Ibba, M. et al. (1997) Science 278:1119-1122).

Alternatively, argininosuccinate synthase activity of ENZM is measured based on the conversion of [^{3}H]aspartate to [^{3}H]argininosuccinate. ENZM is incubated with a mixture of [^{3}H]aspartate, citrulline, Tris-HCl (pH 7.5), ATP, $MgCl_2$, KCl, phosphoenolpyruvate, pyruvate kinase, myokinase, and pyrophosphatase, and allowed to proceed for 60 minutes at 37 °C. Enzyme activity was terminated with addition of acetic acid and heating for 30 minutes at 90 °C. [^{3}H]argininosuccinate is separated from un-catalyzed [^{3}H]aspartate by chromatography and quantified by liquid scintillation spectrometry. The amount of [^{3}H]argininosuccinate detected is proportional to the activity of ENZM in this assay (O'Brien, W. E. (1979) Biochemistry 18:5353-5356).

Alternatively, the esterase activity of ENZM is assayed by the hydrolysis of p-nitrophenylacetate (NPA). ENZM is incubated together with 0.1 μM NPA in 0.1 M potassium phosphate buffer (pH 7.25) containing 150 mM NaCl. The hydrolysis of NPA is measured by the increase of absorbance at 400 nm with a spectrophotometer. The increase in light absorbance is proportional to the activity of ENZM (Probst, M.R. et al. (1994) J. Biol. Chem. 269:21650-21656).

XIX. Identification of ENZM Agonists and Antagonists

Agonists or antagonists of ENZM activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in ENZM activity and antagonists cause a decrease in ENZM activity.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of

embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7511289	1	7511289CD1	39	7511289CB1	
7511056	2	7511056CD1	40	7511056CB1	1443673CA2
7511567	3	7511567CD1	41	7511567CB1	
7511651	4	7511651CD1	42	7511651CB1	
7511881	5	7511881CD1	43	7511881CB1	95003710CA2
7512181	6	7512181CD1	44	7512181CB1	90139166CA2
7511726	7	7511726CD1	45	7511726CB1	
7511057	8	7511057CD1	46	7511057CB1	1449639CA2, 95164104CA2, 95164112CA2, 95164144CA2, 95164160CA2, 95164244CA2, 95164276CA2
7511078	9	7511078CD1	47	7511078CB1	1930540CA2
7511505	10	7511505CD1	48	7511505CB1	
7511552	11	7511552CD1	49	7511552CB1	
7511722	12	7511722CD1	50	7511722CB1	
7511489	13	7511489CD1	51	7511489CB1	
7511497	14	7511497CD1	52	7511497CB1	
7511498	15	7511498CD1	53	7511498CB1	95133328CA2
7511612	16	7511612CD1	54	7511612CB1	6604874CA2
7511624	17	7511624CD1	55	7511624CB1	
7511626	18	7511626CD1	56	7511626CB1	
7512885	19	7512885CD1	57	7512885CB1	90025580CA2
7511965	20	7511965CD1	58	7511965CB1	
7512403	21	7512403CD1	59	7512403CB1	
7512564	22	7512564CD1	60	7512564CB1	95110656CA2
7512646	23	7512646CD1	61	7512646CB1	6883320CA2
7512700	24	7512700CD1	62	7512700CB1	
7512707	25	7512707CD1	63	7512707CB1	95078276CA2
7512710	26	7512710CD1	64	7512710CB1	
7512884	27	7512884CD1	65	7512884CB1	
7512931	28	7512931CD1	66	7512931CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7512933	29	7512933CD1	67	7512933CB1	95137927CA2, 95137951CA2, 95137967CA2, 95138043CA2
7512942	30	7512942CD1	68	7512942CB1	
7513736	31	7513736CD1	69	7513736CB1	90089783CA2, 90089883CA2
7512779	32	7512779CD1	70	7512779CB1	
7512877	33	7512877CD1	71	7512877CB1	90065442CA2
7512782	34	7512782CD1	72	7512782CB1	
7512784	35	7512784CD1	73	7512784CB1	
7512794	36	7512794CD1	74	7512794CB1	8683815CA2
7512886	37	7512886CD1	75	7512886CB1	
7512929	38	7512929CD1	76	7512929CB1	8131571CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7511289CD1	599428|FLJ20481	1.9E-33	[Homo sapiens][Small molecule-binding protein] Protein containing two EF hand domains, which are found in signaling, buffering or transport proteins
2	7511056CD1	g11120435	6.2E-71	[Homo sapiens] alpha-aminoadipic semialdehyde dehydrogenase-phosphopantetheinyl transferase
		569688|AASDHPPT	1.9E-56	[Homo sapiens][Transferase] Member of the 4'-phosphopantetheinyl transferase (holo-acyl carrier protein synthase, ACPS) superfamily
3	7511567CD1	g2725625	2.0E-287	[Homo sapiens] acetolactate synthase
		715550|T26C12.1	1.9E-152	[Caenorhabditis elegans][Lyase][Cytoplasmic] Member of the decarboxylase (tentative) protein family
		609176|Hpcl	3.4E-43	[Mus musculus][Lyase][Cytoplasmic; Peroxisome] Protein with strong similarity to human HPCL2 (peroxisomal 2-hydroxyphytanoyl-CoA lyase), which cleaves 3-methyl-branched fatty acids to make formyl-CoA and 2-methyl-branched fatty aldehydes, member of the thiamine pyrophosphate-dependent enzyme family
4	7511651CD1	g337463	4.1E-123	[Homo sapiens] uroporphyrinogen III synthase (EC 4.2.1.75)
				Tsai, S. F. et al., Human uroporphyrinogen III synthase: molecular cloning, nucleotide sequence, and expression of a full-length cDNA, Proc. Natl. Acad. Sci. U.S.A. 85, 7049-7053 (1988)
		339612|UROS	3.3E-124	[Homo sapiens][Lyase] Uroporphyrinogen III synthase (congenital erythropoietic porphyria), catalyzes the conversion of hydroxymethylbilane to uroporphyrinogen-III in heme biosynthesis; genetic mutation is associated with congenital erythropoietic porphyria
				Deybach, J. C. et al., Point mutations in the uroporphyrinogen III synthase gene in congenital erythropoietic porphyria (Gunther's disease)., Blood 75, 1763-5 (1990).
		581807|Uros	1.1E-95	[Mus musculus][Lyase] Uroporphyrinogen III synthase, catalyzes the conversion of hydroxymethylbilane to uroporphyrinogen-III in heme biosynthesis; mutation of the human UROS gene is associated with congenital erythropoietic porphyria

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Aizencang, G. I. et al. Uroporphyrinogen III synthase. An alternative promoter controls erythroid-specific expression in the murine gene., J Biol Chem 275, 2295-304 (2000).
5	7511881CD1	g1314483	5.4E-73	[Homo sapiens] leukotriene C4 synthase
				Penrose, J. F. et al., Molecular cloning of the gene for human leukotriene C4 synthase. Organization, nucleotide sequence, and chromosomal localization to 5q35, J. Biol. Chem. 271, 11356-11361 (1996)
		344532|LTC4S	4.4E-74	[Homo sapiens][Transferase][Endoplasmic reticulum; Microsomal fraction; Cytoplasmic; Unspecified membrane] Leukotriene C4 synthase, an integral membrane protein that conjugates the epoxide intermediate leukotriene A4 to glutathione, forming the proinflammatory mediator leukotriene C4
				Sjolinder, M. et al., Aberrant expression of active leukotriene C(4) synthase in CD16(+) neutrophils from patients with chronic myeloid leukemia., Blood 95, 1456-64 (2000).
				Hsieh, F. H. et al., T Helper Cell Type 2 Cytokines Coordinately Regulate Immunoglobulin E-dependent Cysteinyl Leukotriene Production by Human Cord Blood-derived Mast Cells. Profound induction of leukotriene c(4) synthase expression by interleukin 4., J Exp Med 193, 123-134. (2001).
		581981|Ltc4s	3.8E-35	[Mus musculus][Transferase] Leukotriene C4 synthase, an integral membrane protein that conjugates the epoxide intermediate leukotriene A4 to glutathione, forming the proinflammatory mediator leukotriene C4; involved in increasing vascular permeability during inflammatory responses
6	7512181CD1	g12751374	2.0E-120	[Homo sapiens] paraoxanase-3
		348104|PON3	3.3E-114	[Homo sapiens][Hydrolase] Paraoxonase 3, member of paraoxonase protein family, which hydrolyze toxic organophosphates, may protect low density lipoprotein against oxidative modification
				Primo-Parmo, S. L. et al., The human serum paraoxonase/arylesterase gene (PON1) is one member of a multigene family., Genomics 33, 498-507 (1996).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		337086|PON2	5.1E-79	[Homo sapiens][Hydrolase] Paraoxonase 2 (arylesterase 2), an antioxidant that protects low density lipoprotein against peroxidation, member of a family of proteins that hydrolyze toxic organophosphates; gene mutations are associated with an increased risk of coronary heart disease
				Sanghera, D. K. et al., DNA polymorphisms in two paraoxonase genes (PON1 and PON2) are associated with the risk of coronary heart disease., Am J Hum Genet 62, 36-44. (1998).
7	7511726CD1	g1922287	2.2E-107	[Homo sapiens] enoyl-CoA hydratase
				Janssen, U. et al., Human mitochondrial enoyl-CoA hydratase gene (ECHS1): structural organization and assignment to chromosome 10q26.2-q26.3, Genomics 40, 470-475 (1997)
		763627|Echs1	1.7E-112	[Rattus norvegicus][Lyase][Cytoplasmic; Mitochondrial] Short chain enoyl-Coenzyme A hydratase 1, catalyzes the hydration of trans-2-enoyl-CoA forming 3-hydroxyacyl-CoA in mitochondrial fatty acid beta-oxidation, a member of the low similarity hydratase/isomerase enzyme superfamily
				Minami-Ishii, N. et al., Molecular cloning and sequence analysis of the cDNA for rat mitochondrial enoyl-CoA hydratase. Structural and evolutionary relationships linked to the bifunctional enzyme of the peroxisomal beta- oxidation system., Eur J Biochem 185, 73-8 (1989).
				Kiema, T. R. et al., Mutagenic and enzymological studies of the hydratase and isomerase activities of 2-enoyl-CoA hydratase-1., Biochemistry 38, 2991-9. (1999).
		662412|ECHS1	1.8E-108	[Homo sapiens][Lyase][Cytoplasmic; Mitochondrial] Short chain enoyl-Coenzyme A hydratase 1, catalyzes the second step in mitochondrial fatty acid beta-oxidation
				Sakata, M. et al., Messenger RNA differential display reverse-transcriptase-polymerase-chain-reaction analysis of a progestogen-suppressive gene in a human endometrial-cancer cell line., Int J Cancer 78, 125-9 (1998).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		732665|Echs1	2.1E-89	[Mus musculus] Protein with strong similarity to short chain enoyl-Coenzyme A hydratase 1 (rat Echs1), which catalyzes the hydration of trans-2-enoyl-CoA in mitochondrial fatty acid beta-oxidation, member of the enoyl-CoA hydratase or isomerase family
8	7511057CD1	g7159801	2.8E-59	[Homo sapiens] dJ351K20.2.1 (novel enoyl coA/acyl coA hydratase/dehydrogenase type protein (isoform 1))
		599822|LOC55862	1.2E-56	[Homo sapiens][Lyase; Oxidoreductase] Uncharacterized hypothalamus protein HCDASE, a member of the enoyl-CoA hydratase/isomerase family, which function in fatty acid metabolism
9	7511078CD1	g18478646	3.1E-168	[Homo sapiens] p53-responsive gene 3
		731657|FLJ14497	2.5E-169	[Homo sapiens] Member of the pyridine nucleotide-disulfide oxidoreductase family of flavoproteins, has weak similarity to C. albicans orf6.6088
		641934|orf6.6088	2.1E-18	[Candida albicans][Oxidoreductase] Member of the pyridine nucleotide-disulphide oxidoreductase family of flavoproteins, has low similarity to uncharacterized S. cerevisiae Ynr074p
				Huang, S. et al., Specificity of cotranslational amino-terminal processing of proteins in yeast., Biochemistry 26, 8242-6 (1987).
10	7511505CD1	g4106362	1.4E-251	[Homo sapiens] 7-dehydrocholesterol reductase
				Waterham, H. R. et al., Smith-Lemli-Opitz syndrome is caused by mutations in the 7-dehydrocholesterol reductase gene, Am. J. Hum. Genet. 63, 329-338 (1998)
				Yu, H. et al., Spectrum of Delta(7)-dehydrocholesterol reductase mutations in patients with the smith-lemli-opitz (RSH) syndrome, Hum. Mol. Genet. 9, 1385-1391 (2000)
		335038|DHCR7	5.0E-252	[Homo sapiens][Oxidoreductase][Endoplasmic reticulum; Cytoplasmic; Unspecified membrane] 7-dehydrocholesterol reductase, catalyzes the reduction of the C7-C8 (delta 7) double bond of 7-dehydrocholesterol in the last step of cholesterol biosynthesis; mutations in the corresponding gene are associated with Smith-Lemli-Opitz syndrome

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Kim, J. H. et al., Cholesterol biosynthesis from lanosterol. A concerted role for Sp1 and NF-Y-binding sites for sterol-mediated regulation of rat 7-dehydrocholesterol reductase gene expression., J Biol Chem 276, 18153-60. (2001).
		583989|Dhcr7	1.7E-227	[Mus musculus][Oxidoreductase] 7-dehydrocholesterol reductase, catalyzes the reduction of the C7-C8 (delta 7) double bond of 7-dehydrocholesterol in the last step of cholesterol biosynthesis; mutations in the human DHCR7 gene are associated with Smith-Lemli-Opitz syndrome
				Wassif, C. A. et al., Biochemical, phenotypic and neurophysiological characterization of a genetic mouse model of RSH/Smith--Lemli--Opitz syndrome., Hum Mol Genet 10, 555-64. (2001).
				Kim, J. H. et al. (supra)
11	7511552CD1	g178372	3.3E-113	[Homo sapiens] aldehyde dehydrogenase
				Hsu, L. C. et al., Genomic structure of the human cytosolic aldehyde dehydrogenase gene, Genomics 5, 857-865 (1989)
		334128|ALDH1A1	8.9E-114	[Homo sapiens][Oxidoreductase; Small molecule-binding protein][Cytoplasmic] Aldehyde dehydrogenase 1 family member A1, a broad specificity cytosolic NAD-dependent enzyme that oxidizes aldehydes to carboxylic acids, oxidizes retinal to retinoic acid, detoxifies xenobiotics, and contributes to cyclophosphamide resistance
				Magni, M. et al., Induction of cyclophosphamide-resistance by aldehyde-dehydrogenase gene transfer., Blood 87, 1097-103. (1996).
		586863|Aldh1a1	2.4E-95	[Mus musculus][Oxidoreductase][Cytoplasmic] Aldehyde dehydrogenase 1 family member A1, a broad specificity cytosolic NAD-dependent enzyme that oxidizes aldehydes to carboxylic acids, oxidizes retinal to retinoic acid, detoxifies xenobiotics, and contributes to cyclophosphamide resistance
				Greene, W. K. et al., The T-cell oncogenic protein HOX11 activates Aldh1 expression in NIH 3T3 cells but represses its expression in mouse spleen development., Mol Cell Biol 18, 7030-7 (1998).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
12	7511722CD1	g17985315	4.7E-41	[Brucella melitensis] HYDROXYPYRUVATE ISOMERASE
				DelVecchio, V. G. et al., The genome sequence of the facultative intracellular pathogen Brucella melitensis, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448 (2002)
		238614|C05D11.5	1.6E-34	[Caenorhabditis elegans][Unknown] Protein with strong similarity to D. melanogaster GIP glyoxylate-induced protein
				Reinke, V. et al., A global profile of germline gene expression in C. elegans., Mol Cell 6, 605-16 (2000).
13	7511489CD1	g307432	3.2E-190	[Homo sapiens] squalene synthetase
				McKenzie, T. L. et al., Molecular cloning, expression, and characterization of the cDNA for the rat hepatic squalene synthase, J. Biol. Chem. 267, 21368-21374 (1992)
				Jiang, G. et al., Transcriptional regulation by lovastatin and 25-hydroxycholesterol in HepG2 cells and molecular cloning and expression of the cDNA for the human hepatic squalene synthase, J. Biol. Chem. 268, 12818-12824 (1993)
		340542|FDFT1	2.6E-191	[Homo sapiens][Transferase] Squalene synthase (farnesyl-diphosphate farnesyltransferase 1), catalyzes the conversion of farnesyl diphosphate to squalene in cholesterol biosynthesis, may be a potential target for cholesterol lowering therapy
				Tozawa, R. et al., Embryonic lethality and defective neural tube closure in mice lacking squalene synthase., J Biol Chem 274, 30843-8 (1999).
		584775|Fdft1	3.1E-171	[Mus musculus][Transferase] Squalene synthase (farnesyl-diphosphate farnesyltransferase 1), catalyzes the conversion of farnesyl diphosphate to squalene in cholesterol biosynthesis
				Sakakura, Y. et al., Sterol regulatory element-binding proteins induce an entire pathway of cholesterol synthesis., Biochem Biophys Res Commun 286, 176-83. (2001).
14	7511497CD1	g2911587	1.2E-44	[Homo sapiens] methyltransferase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Stanchi, F. et al., Characterization of 16 novel human genes showing high similarity to yeast sequences, Yeast 18, 69-80 (2001)
		240044|C27F2.4	3.6E-23	[Caenorhabditis elegans][Unknown] Putative ortholog of S. cerevisiae Ycr047p, a protein with similarity to rat methylglycine transferase
				Wang, S. P. et al., Structure-function analysis of the mRNA cap methyltransferase of Saccharomyces cerevisiae., J Biol Chem 272, 14683-9 (1997).
		2051|BUD23	9.7E-21	[Saccharomyces cerevisiae][Transferase] Protein with similarity to rat methylglycine transferase, may be involved in bud site selection/bipolar budding
				Moerschell, R. P. et al., The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes c created by oligonucleotide transformation., J Biol Chem 265, 19638-43 (1990).
15	7511498CD1	g16226067	2.2E-114	[Homo sapiens] putative methyltransferase WBMT
		240044|C27F2.4	4.0E-63	[Caenorhabditis elegans][Unknown] Putative ortholog of S. cerevisiae Ycr047p, a protein with similarity to rat methylglycine transferase
				Wang, S. P. et al. (supra)
		2051|BUD23	5.8E-55	[Saccharomyces cerevisiae][Transferase] Protein with similarity to rat methylglycine transferase, may be involved in bud site selection/bipolar budding
				Slonimski, P. P. et al. (supra)
				Moerschell, R. P. et al. (supra)
16	7511612CD1	g12653701	3.2E-67	[Homo sapiens] acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
		333998|ACAA1	2.6E-68	[Homo sapiens][Transferase][Cytoplasmic; Peroxisome] Peroxisomal 3-oxoacyl-coenzyme A thiolase (peroxisomal 3-ketoacyl-Coenzyme A thiolase), functions in peroxisomal beta-oxidation of fatty acids, deficiency is associated with rhizomelic chondrodysplasia punctata (RCDP)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Heikoop, J. C. et al., Rhizomelic chondrodysplasia punctata. Deficiency of 3-oxoacyl-coenzyme A thiolase in peroxisomes and impaired processing of the enzyme., J Clin Invest 86, 126-30. (1990).
		430496|Acaa	2.9E-51	[Rattus norvegicus][Transferase][Cytoplasmic; Peroxisome] Peroxisomal 3-ketoacyl-CoA thiolase A (1), catalyzes the final step in fatty acid beta-oxidation, expression is only weakly activated by administration of peroxisome proliferators
				Hijikata, M. et al., Rat peroxisomal 3-ketoacyl-CoA thiolase gene. Occurrence of two closely related but differentially regulated genes., J Biol Chem 265, 4600-6 (1990).
17	7511624CD1	g179104	1.6E-199	[Homo sapiens] aspartate aminotransferase precursor (2.6.1.1)
				Pol, S. et al., Nucleotide sequence and tissue distribution of the human mitochondrial aspartate aminotransferase mRNA, Biochem. Biophys. Res. Commun. 157, 1309-1315 (1988)
		343324|GOT2	1.3E-200	[Homo sapiens][Transferase][Cytoplasmic; Mitochondrial] Mitochondrial aspartate aminotransferase (glutamic oxaloacetic transaminase), catalyzes the reversible transfer of the amino group from aspartate to 2-oxoglutarate to form oxaloacetate and glutamate
				Salomon, A. R. et al., Understanding and exploiting the mechanistic basis for selectivity of polyketide inhibitors of F0F1-ATPase, Proc Natl Acad Sci U S A 97, 14766-71 (2000).
		584895|Got2	1.1E-192	[Mus musculus][Transferase][Cytoplasmic; Mitochondrial] Mitochondrial aspartate aminotransferase, catalyzes the reversible transfer of the amino group from aspartate to 2-oxoglutarate to form oxaloacetate and glutamate
				Memon, R. A. et al., Regulation of putative fatty acid transporters and Acyl-CoA synthetase in liver and adipose tissue in ob/ob mice., Diabetes 48, 121-7 (1999).
18	7511626CD1	g179104	1.1E-209	[Homo sapiens] aspartate aminotransferase precursor (2.6.1.1)
				Pol, S. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		343324|GOT2	8.9E-211	[Homo sapiens][Transferase][Cytoplasmic; Mitochondrial] Mitochondrial aspartate aminotransferase (glutamic oxaloacetic transaminase), catalyzes the reversible transfer of the amino group from aspartate to 2-oxoglutarate to form oxaloacetate and glutamate
				Salomon, A. R. et al. (supra)
		584895|Got2	4.2E-202	[Mus musculus][Transferase][Cytoplasmic; Mitochondrial] Mitochondrial aspartate aminotransferase, catalyzes the reversible transfer of the amino group from aspartate to 2-oxoglutarate to form oxaloacetate and glutamate
				Memon, R. A. et al. (supra)
19	7512885CD1	569696| DKFZp566O084	6.1E-138	[Homo sapiens][Oxidoreductase] Member of the short-chain dehydrogenase/reductase family
		251240|T25G12.2	7.0E-28	[Caenorhabditis elegans][Oxidoreductase] Putative dehydrogenase, has strong similarity over the N-terminal region to C. elegans T25G12.7
				Bateman, A. et al., Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins., Nucleic Acids Res 27, 260-2 (1999).
		754109|Hsd17b11	2.7E-18	[Mus musculus] Member of the short-chain dehydrogenase-reductase family, which are NAD- or NADP-dependent oxidoreductases, has low similarity to short-chain dehydrogenase-reductase 1 (human SDR1), which catalyzes the reduction of all-trans-retinal to retinol
20	7511965CD1	g3150035	1.9E-146	[Homo sapiens] aldose reductase-like peptide
				Cao, D. et al., Identification and characterization of a novel human aldose reductase-like gene, J. Biol. Chem. 273, 11429-11435 (1998)
		690816|AKR1B10	5.0E-147	[Homo sapiens][Oxidoreductase] Aldose reductase-like protein (human small intestine reductase), reduces aliphatic and aromatic aldehydes and, to a lesser extent, hexoses, may detoxify reactive digested aldehydes or act in steroid metabolism, overexpressed in some liver cancers

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Lefrancois-Martinez, A. M. et al., Product of side-chain cleavage of cholesterol, isocaproaldehyde, is an endogenous specific substrate of mouse vas deferens protein, an aldose reductase-like protein in adrenocortical cells., J Biol Chem 274, 32875-80 (1999).
	7511965CD1	589811|Akr1b1	9.6E-110	[Rattus norvegicus][Oxidoreductase] Aldose reductase, member of the NADPH-dependent aldo-keto reductase superfamily, reduces glucose and other carbonyl-containing substrates as part of the polyol pathway, metabolizes xenobiotics, and may contribute to diabetic complications
				Takahashi, M. et al., Elevation of aldose reductase gene expression in rat primary hepatoma and hepatoma cell lines: implication in detoxification of cytotoxic aldehydes., Int J Cancer 62, 749-54 (1995).
21	7512403CD1	g1870244	7.1E-197	[Homo sapiens] aldehyde dehydrogenase
				Chang, C. et al., Human fatty aldehyde dehydrogenase gene (ALDH10): organization and tissue-dependent expression, Genomics 40, 80-85 (1997)
		339660|ALDH3A2	5.6E-198	[Homo sapiens][Oxidoreductase] Aldehyde dehyrogenase 3 family member A2, catalyzes the oxidation of long chain fatty aldehydes and leukotrienes; mutation of the corresponding gene causes Sjogren Larsson syndrome, a disorder marked by mental retardation, spasticity, and ichthyosis
				Kraus, C. et al., RNA-based mutation screening in German families with Sjogren-Larsson syndrome., Eur J Hum Genet 8, 299-306. (2000).
		711692|Aldh3a2	6.7E-167	[Rattus norvegicus][Oxidoreductase][Endoplasmic reticulum; Cytoplasmic; Plasma membrane] Aldehyde dehydrogenase family 3 subfamily A2, catalyzes the oxidation of long and medium chain aliphatic aldehydes derived from lipid metabolism; mutation of the human ALDH3A2 gene causes Sjogren Larsson syndrome
				Rizzo, W. B. et al., Fatty aldehyde dehydrogenase: genomic structure, expression and mutation analysis in Sjogren-Larsson syndrome., Chem Biol Interact 130, 297-307. (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
22	7512564CD1	g9622124	6.7E-76	[Homo sapiens] androgen-regulated short-chain dehydrogenase/reductase 1
				Lin, B. et al., Prostate short-chain dehydrogenase reductase 1 (PSDR1): a new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium, Cancer Res. 61, 1611-1618 (2001)
		475723|LOC51109	5.2E-77	[Homo sapiens] Protein containing EGF-like domains, which are found in some secreted proteins and extracellular domains of transmembrane proteins
		613361|PAN2	7.0E-42	[Homo sapiens] Member of the short-chain dehydrogenase-reductase family, which are NAD- or NADP-dependent oxidoreductases, has low similarity to 17 beta-hydroxysteroid dehydrogenase type 7 (mouse Hsd17b7), which converts estrone to estradiol in the corpus luteum
23	7512646CD1	g18203832	2.7E-95	[Homo sapiens] peroxiredoxin 3
		428166|PRDX3	2.1E-96	[Homo sapiens][Oxidoreductase] Peroxiredoxin 3, a mitochondrial antioxidant protein involved in oxygen and radical metabolism, has an antiapoptotic function
				Shih, S. F. et al., Abrin triggers cell death by inactivating a thiol-specific antioxidant protein., J Biol Chem 276, 21870-7. (2001).
		583587|Prdx3	1.1E-78	[Mus musculus] Peroxiredoxin 3, a putative antioxidant protein, has alkyl hydroperoxide reductase activity, involved in cell differentiation
				Tsuji, K. et al., Mammalian antioxidant protein complements alkylhydroperoxide reductase (ahpC) mutation in Escherichia coli., Biochem J 307, 377-81 (1995).
24	7512700CD1	g431857	4.0E-144	[Homo sapiens] delta 4-3-oxosteroid 5 beta-reductase
				Kondo, K. H. et al., Cloning and expression of cDNA of human delta 4-3-oxosteroid 5 beta-reductase and substrate specificity of the expressed enzyme, Eur. J. Biochem. 219, 357-363 (1994)
		343204|AKR1D1	3.1E-145	[Homo sapiens][Oxidoreductase] Delta 4-3-oxosteroid 5 beta-reductase, catalyzes the reduction of the Delta 4 double bond of bile acid intermediates and steroid hormones; deficiency results in neonatal liver failure with associated hemochromatosis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Shneider, B. L. et al., Delta 4-3-oxosteroid 5 beta-reductase deficiency causing neonatal liver failure and hemochromatosis., J Pediatr 124, 234-8. (1994).
		332522|Rn.25716	5.6E-118	[Rattus norvegicus][Oxidoreductase] Delta 4-3-oxosteroid 5 beta-reductase, catalyzes the reduction of the Delta 4 double bond of bile acid intermediates and steroid hormones; deficiency of human AKR1D1 results in neonatal liver failure with associated hemochromatosis
				Kondo, K. H. et al., Cloning and expression of cDNA of human delta 4-3-oxosteroid 5 beta-reductase and substrate specificity of the expressed enzyme., Eur J Biochem 219, 357-63 (1994).
25	7512707CD1	g9951529	1.9E-32	[Pseudomonas aeruginosa] Probable FAD-dependent monooxygenase
		475553|LOC51004	3.4E-153	[Homo sapiens] Member of the flavoprotein monooxygenase family of aromatic ring hydroxylases, has moderate similarity to uncharacterized C. elegans K07B1.2
		247658|K07B1.2	6.2E-42	[Caenorhabditis elegans][Oxidoreductase][Mitochondrial] Protein with strong similarity to S. cerevisiae Coq6p, a monooxygenase required for coenzyme Q (ubiquinone) biosynthesis
				Hill, A. A. et al., Genomic analysis of gene expression in C. elegans., Science 290, 809-12. (2000).
26	7512710CD1	475553|LOC51004	1.5E-217	[Homo sapiens] Member of the flavoprotein monooxygenase family of aromatic ring hydroxylases, has moderate similarity to uncharacterized C. elegans K07B1.2
		7170|COQ6	4.5E-54	[Saccharomyces cerevisiae][Oxidoreductase][Mitochondrial] Monooxygenase required for coenzyme Q (ubiquinone) biosynthesis
				Hsu, A. Y. et al., Genetic evidence for a multi-subunit complex in the O-methyltransferase steps of coenzyme Q biosynthesis., Biochim Biophys Acta 1484, 287-297 (2000).
27	7512884CD1	g1747331	7.9E-41	[Homo sapiens] cysteine dioxygenase
		334614|CDO1	2.7E-41	[Homo sapiens][Oxidoreductase] Cysteine dioxygenase 1, catalyzes the conversion of cysteine to 3-sulfinoalanine; the primary enzyme in cysteine catabolism and the regulation of sulfate production

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Tsuboyama-Kasaoka, N. et al., Human cysteine dioxygenase gene: structural organization, tissue-specific expression and downregulation by phorbol 12-myristate 13-acetate., Biosci Biotechnol Biochem 63, 1017-24. (1999).
		751686|Cdo1	4.5E-39	[Rattus norvegicus][Oxidoreductase] Cysteine dioxygenase 1, catalyzes the conversion of cysteine to 3-sulfinoalanine in cysteine catabolism
				Parsons, R. B. et al., Cysteine dioxygenase: regional expression of activity in rat brain., Neurosci Lett 248, 101-4. (1998).
28	7512931CD1	g837328	1.1E-242	[Homo sapiens] protoporphyrinogen oxidase
				Dailey, T. A. et al., Expression of a cloned protoporphyrinogen oxidase, J. Biol. Chem. 269, 813-815 (1994)
		337132|PPOX	8.9E-244	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX in heme biosynthesis, inhibited by the tetrahydrophthalimide and diphenyl ether herbicides; genetic mutation is detected in patients with variegate porphyria
				Whatley, S. D. et al., Variegate porphyria in Western Europe: identification of PPOX gene mutations in 104 families, extent of allelic heterogeneity, and absence of correlation between phenotype and type of mutation., Am J Hum Genet 65, 984-94. (1999).
		582571|Ppox	2.5E-215	[Mus musculus][Oxidoreductase] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, inhibited by agents used in light treatment of tumors and by diphenyl ether herbicides; deficiency of human PPOX is associated with variegate porphyria
				Fingar, V. H. et al., Photodynamic therapy using a protoporphyrinogen oxidase inhibitor., Cancer Res 57, 4551-6. (1997).
29	7512933CD1	g837328	3.5E-94	[Homo sapiens] protoporphyrinogen oxidase
				Dailey, T. A. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		337132|PPOX	2.8E-95	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX in heme biosynthesis, inhibited by the tetrahydrophthalimide and diphenyl ether herbicides; genetic mutation is detected in patients with variegate porphyria
				Whatley, S. D. et al. (supra)
		582571|Ppox	6.9E-82	[Mus musculus][Oxidoreductase] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, inhibited by agents used in light treatment of tumors and by diphenyl ether herbicides; deficiency of human PPOX is associated with variegate porphyria
				Fingar, V. H. et al. (supra)
30	7512942CD1	g837328	9.1E-56	[Homo sapiens] protoporphyrinogen oxidase
				Dailey, T. A. et al. (supra)
		337132|PPOX	7.1E-57	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX in heme biosynthesis, inhibited by the tetrahydrophthalimide and diphenyl ether herbicides; genetic mutation is detected in patients with variegate porphyria
				Whatley, S. D. et al. (supra)
		582571|Ppox	5.7E-48	[Mus musculus][Oxidoreductase] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, inhibited by agents used in light treatment of tumors and by diphenyl ether herbicides; deficiency of human PPOX is associated with variegate porphyria
				Fingar, V. H. et al. (supra)
31	7513736CD1	g7208438	8.1E-146	[Homo sapiens] long-chain 2-hydroxy acid oxidase HAOX2
				Jones, J. M. et al., Identification of HAOX1, HAOX2, and HAOX3, Three Human Persoxisomal 2-Hydroxy Acid Oxidases, J. Biol. Chem. (2000) 275(17):12590-12597

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		475383|HAOX2	6.4E-147	[Homo sapiens][Oxidoreductase][Cytoplasmic; Peroxisome] Hydroxy acid oxidase 2, oxidizes 2-hydroxy fatty acids and displays highest activity toward 2 hydroxypalmitate, may contribute to the general fatty acid alpha-oxidation pathway
		747907|Hao3	7.7E-108	[Rattus norvegicus][Oxidoreductase] Hydroxyacid oxidase 3 (long-chain L-alpha-hydroxy acid oxidase), a member of the FMN-dependent alpha-hydroxy acid-oxidizing family, may function in lipid metabolism
				Belmouden, A. et al., Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase. Expression of the catalytically active recombinant protein as a chimaera., Eur J Biochem 214, 17-25 (1993).
32	7512779CD1	g4894560	4.7E-217	[Homo sapiens] cysteine sulfinic acid decarboxylase-related protein 3
		624468|Csad	6.8E-203	[Rattus norvegicus][Lyase] Cysteine sulfinic acid decarboxylase, an enzyme involved in the biosynthesis of taurine, expression is repressed by triiodothyronine and estrogen, and expression is stimulated during hepatocarcinogenesis
				Kishimoto, T. et al., Overexpression of cysteine sulfinic acid decarboxylase stimulated by hepatocarcinogenesis results in autoantibody production in rats., Cancer Res 56, 5230-7 (1996).
		335440|GAD2	7.0E-114	[Homo sapiens][Lyase] Glutamate decarboxylase (L-glutamate 1-carboxy-lyase), converts L glutamic acid to gamma-aminobutyric acid; an autoantigen in insulin-dependent diabetes mellitus, and stiff-man syndrome; murine Gad2 is associated with epilepsy
				Mally, M. I. et al., Ontogeny and tissue distribution of human GAD expression., Diabetes 45, 496-501 (1996).
33	7512877CD1	g1483511	1.1E-43	[Homo sapiens] 3-hydroxyacyl-CoA dehydrogenase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Vredendaal, P. J. et al., Human short-chain L-3-hydroxyacyl-CoA dehydrogenase: cloning and characterization of the coding sequence, Biochem. Biophys. Res. Commun. 223, 718-723 (1996)
		342104|HADHSC	8.6E-45	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Short chain L-3-hydroxyacyl-Coenzyme A, a homodimeric enzyme that catalyzes the penultimate step in the fatty acid beta oxidation spiral, deficiency results in hyperinsulinism with elevated blood hydroxybutyrylcarnitine
				Clayton, P. T. et al., Hyperinsulinism in short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency reveals the importance of beta-oxidation in insulin secretion., J Clin Invest 108, 457-65. (2001).
		726175|1f12_A	6.0E-44	[Protein Data Bank] L-3-Hydroxyacyl-Coa Dehydrogenase
34	7512782CD1	g1060907	1.3E-54	[Homo sapiens] quinolinate phosphoribosyl transferase
				Fukuoka, S. I. et al., Characterization and functional expression of the cDNA encoding human brain quinolinate phosphoribosyltransferase, Biochim. Biophys. Acta 1395, 192-201 (1998)
		569398|QPRT	1.0E-55	[Homo sapiens][Transferase] Quinolinate phosphoribosyltransferase, catabolizes quinolinate, an intermediate in NAD synthesis and a neuron exitotoxin; excess expression is associated with neurodegenerative disorders and decreased activity may play a role in epilepsy
				Feldblum, S. et al., Quinolinic-phosphoribosyl transferase activity is decreased in epileptic human brain tissue., Epilepsia 29, 523-9. (1988).
		9762|YFR047C	1.7E-31	[Saccharomyces cerevisiae][Transferase] Putative quinolinic acid phosphoribosyltransferase, active in the pathway of tryptophan degradation and nicotinic acid synthesis
35	7512784CD1	g1060907	2.5E-53	[Homo sapiens] quinolinate phosphoribosyl transferase
				Fukuoka, S. I. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		569398|QPRT	2.0E-54	[Homo sapiens][Transferase] Quinolinate phosphoribosyltransferase, catabolizes quinolinate, an intermediate in NAD synthesis and a neuron exitotoxin; excess expression is associated with neurodegenerative disorders and decreased activity may play a role in epilepsy
				Feldblum, S. et al. (supra)
		9762|YFR047C	3.6E-24	[Saccharomyces cerevisiae][Transferase] Putative quinolinic acid phosphoribosyltransferase, active in the pathway of tryptophan degradation and nicotinic acid synthesis
36	7512794CD1	g178518	1.9E-15	[Homo sapiens] S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) old gene name 'AMD'
				Pajunen, A. et al., Structure and regulation of mammalian S-adenosylmethionine decarboxylase, J. Biol. Chem. 263, 17040-17049 (1988)
		343256|AMD1	1.5E-16	[Homo sapiens][Lyase] S-adenosylmethionine decarboxylase 1, catalyzes the removal of the carboxylate group of S-adenosylmethionine to form S-adenosyl-5'-3-methylpropylamine in the polyamine biosynthesis pathway
				Bettuzzi, S. et al., Tumor progression is accompanied by significant changes in the levels of expression of polyamine metabolism regulatory genes and clusterin (sulfated glycoprotein 2) in human prostate cancer specimens [published erratum appears in Cancer Res 2000 Mar 1; 60(5):1472], Cancer Res 60, 28-34 (2000).
37	7512886CD1	g4239986	9.0E-88	[Homo sapiens] PIG-L
				Watanabe, R. et al., Mammalian PIG-L and its yeast homologue Gpi12p are N-acetylglucosaminylphosphatidylinositol de-N-acetylases essential in glycosylphosphatidylinositol biosynthesis, Biochem. J. 339 (Pt 1), 185-192 (1999)
		341062|PIGL	7.1E-89	[Homo sapiens][Hydrolase] Protein with high similarity to N-acetylglucosaminylphosphatidylinositol de-N-acetylase (rat Rn.10774), which catalyzes the second step in GPI anchor synthesis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Smith, T. K. et al., Specificity of GlcNAc-PI de-N-acetylase of GPI biosynthesis and synthesis of parasite-specific suicide substrate inhibitors., Embo Journal 20, 3322-32. (2001).
		331106|Pigl	1.0E-64	[Rattus norvegicus][Hydrolase] N-acetylglucosaminylphosphatidylinositol de-N-acetylase, catalyzes the N-deacylation of N-acetylglucosaminyl phosphatidylinositol (GlcNAc-PI) to form glucosaminyl phosphatidylinositol (GlcN-PI) in the second step of GPI anchor biosynthesis
				Nakamura, N. et al., Expression cloning of PIG-L, a candidate N-acetylglucosaminyl-phosphatidylinositol deacetylase., J Biol Chem 272, 15834-40 (1997).
38	7512929CD1	g15150358	1.0E-40	[Mus musculus] UDP-glucuronic acid decarboxylase

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7511289CD1	256	S5 S148 S171 S177 S207 T131 T190	N152	Signal_cleavage: M1-L67	SPSCAN
					Signal Peptide: L46-L67	HMMER
					AGP_acyltrn: 1-acyl-sn-glycerol-3-phosphate acyltransferases: L107-S237	HMMER_TIGRFAM
					Cytosolic domain: M1-F39; Transmembrane domain: C40-W62; Non-cytosolic domain: P63-P256	TMHMMER
					Leucine zipper pattern: L46-L67	MOTIFS
2	7511056CD1	138	S29 T87	N108	Signal_cleavage: M1-A31	SPSCAN
					PROTEIN BIOSYNTHESIS ANTIBIOTIC SURFACTIN LIPOPEPTIDE ANTIBIOTICS ITURIN A GSP PRODUCTION PD005299: I40-I130	BLAST_PRODOM
3	7511567CD1	674	S50 S178 S367 S527 S571 S638 S648 S671 T95 T188 T299 T352 T432 T457	N336	Signal Peptide: M1-A28	HMMER
					Thiamine pyrophosphate enzyme, central: S269-L422	HMMER_PFAM
					Thiamine pyrophosphate enzyme, N-terminus: S50-K221	HMMER_PFAM
					menD: 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase: V46-F169	HMMER_TIGRFAM
					Cytosolic domain: M1-G11; Transmembrane domain: S12-H34; Non-cytosolic domain: R35-D674	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Thiamine pyrophosphate dependent enzyme IPB000399: L62-C85, A105-A122, P330-S359, G494-A506, P511-G548	BLIMPS_BLOCKS
					Thiamine pyrophosphate enzymes signature: P484-R542	PROFILESCAN
					THIAMINE PYROPHOSPHATE FLAVOPROTEIN SYNTHASE LYASE PYRUVATE ACETOLACTATE DECARBOXYLASE ACID SUBUNIT PD000376: L62-I539	BLAST_PRODOM
					ACETOLACTATE SYNTHASE FLAVOPROTEIN LYASE THIAMINE PYROPHOSPHATE HOMOLOG PD039458: M1-V61	BLAST_PRODOM
					THIAMINE PYROPHOSPHATE ENZYMES DM00391|P08142|10-559: G54-K230, E254-T533, P9-G37	BLAST_DOMO
					THIAMINE PYROPHOSPHATE ENZYMES DM00391|P39994|1-559: E56-L213, P270-R536	BLAST_DOMO
					THIAMINE PYROPHOSPHATE ENZYMES DM00391|P00893|1-563: G54-P215 Q272-I539 R561-R578	BLAST_DOMO
					THIAMINE PYROPHOSPHATE ENZYMES DM00391|P40149|7-564: L62-P209 Q256-I539	BLAST_DOMO
					Thiamine pyrophosphate enzymes signature: L504-A523	MOTIFS
4	7511651CD1	237	S23 S35 S58 S176 T86 T200	N50 N73	Signal_cleavage: M1-A41	SPSCAN
					Uroporphyrinogen-III synthase HemD: K2-G226	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					UROPORPHYRINOGENIII SYNTHASE UROS UROPORPHYRINOGEN III COSYNTHETASE HYDROXYMETHYLBILANE HYDROLYASE CYCLIZING UROIIIS PD017458: M1-R21 S23-I227	BLAST_PRODOM
					UROPORPHYRINOGEN-III SYNTHASE DM07953|P10746|1-264: M1-R21 R21-C237	BLAST_DOMO
					UROPORPHYRINOGEN-III SYNTHASE DM07953|P51163|1-264: M1-R21 R21-C237	BLAST_DOMO
					Aldehyde dehydrogenases cysteine active site: L84-E95	MOTIFS
5	7511881CD1	178	S28 S139	N55	Signal_cleavage: M1-L18	SPSCAN
					Signal Peptide: M1-A20	HMMER
					MAPEG family: P44-A161	HMMER_PFAM
					Cytosolic domains: Q25-S57, Q123-A142; Transmembrane domains: V5-L24, E58-V80, C100-F122, L143-A165; Non-cytosolic domains: M1-E4, G81-P99, A166-A178	TMHMMER
					MAPEG (Membrane-associated proteins in eicosanoid and glutathione metabolism) IPB001129: G42-G79	BLIMPS_BLOCKS
					TRANSMEMBRANE LEUKOTRIENE BIOSYNTHESIS PROTEIN TRANSFERASE 5LIPOXYGENASE ACTIVATING FLAP MK886BINDING MICROSOMAL PD009387: M1-H75 A107-T173	BLAST_PRODOM
					FLAP/GST2/LTC4S family signature: G42-C56	MOTIFS
6	7512181CD1	245	S50 S138 S165 S191 S208 T157	N29	Signal_cleavage: M1-V17	SPSCAN
					Signal Peptide: M1-S15	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M1-G18	HMMER
					Signal Peptide: M1-A23	HMMER
					Arylesterase: G2-G245	HMMER_PFAM
					Cytosolic domain: M1-L4; Transmembrane domain: V5-F24; Non-cytosolic domain: R25-G245	TMHMMER
					SERUM PARAOXONASE/ARYLESTERASE P. PD02637: R32-F86, F106-V140, E141-I169, V170-K217, G218-G245	BLIMPS_PRODOM
					SERUM PARAOXONASE/ARYLESTERASE PON ARYLDIAKYLPHOSPHATASE AESTERASE AROMATIC ESTERASE HYDROLASE GLYCOPROTEIN Signal PD005046: E49-Y233	BLAST_PRODOM
					PARAOXONASE/ARYLESTERASE SERUM PON ARYLDIAKYLPHOSPHATASE AESTERASE AROMATIC ESTERASE HYDROLASE GLYCOPROTEIN Signal PD005529: M1-E46	BLAST_PRODOM
					PARAOXONASE; ARYLESTERASE; SERUM; DM07178|P54832|1-353: M1-Y233	BLAST_DOMO
					PARAOXONASE; ARYLESTERASE; SERUM; DM07178|P27169|1-353: K3-Y233	BLAST_DOMO
7	7511726CD1	250	S107 S113 S199 S216 T88 T176 T211	N44 N105	Enoyl-CoA hydratase/isomerase family: G48-E214	HMMER_PFAM
					Enoyl-CoA hydratase/isomerase IPB001753: I50-L61, G83-N105, K128-A154, G175-E214	BLIMPS_BLOCKS
					Enoyl-CoA hydratase/isomerase signature: K118-T169	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN HYDRATASE ENOYLCOA ACID FATTY LYASE ISOMERASE METABOLISM 3HYDROXYACYLCOA DEHYDROGENASE PD000432: V47-A203	BLAST_PRODOM
					HYDRATASE ENOYLCOA MITOCHONDRIAL PRECURSOR SHORT CHAIN SCEH FATTY ACID METABOLISM PD000523: M1-G48	BLAST_PRODOM
					ENOYL-COA HYDRATASE/ISOMERASE DM00366|P30084|34-285: E34-A206, A200-N246	BLAST_DOMO
					ENOYL-COA HYDRATASE/ISOMERASE DM00366|P34559|32-283: E34-Q205	BLAST_DOMO
					ENOYL-COA HYDRATASE/ISOMERASE DM00366|P52046|1-255: I36-L218	BLAST_DOMO
					ENOYL-COA HYDRATASE/ISOMERASE DM00366|I37195|76-334: K43-K204, L187-R243	BLAST_DOMO
					Enoyl-CoA hydratase/isomerase signature: I131-I151	MOTIFS
8	7511057CD1	145	S12 S107	N89	Enoyl-CoA hydratase/isomerase IPB001753: L61-F72, K94-S116	BLIMPS_BLOCKS
9	7511078CD1	333	S147 S166 S170 S272 S324 T72 T188 T251 T327	N208	Signal_cleavage: M1-A26	SPSCAN
					Pyridine nucleotide-disulphide oxidoreductase: H13-Y256	HMMER_PFAM
					Cytosolic domain: M1-G10; Transmembrane domain: A11-L33; Non-cytosolic domain: N34-P333	TMHMMER
					FAD-dependent pyridine nucleotide reductase signature PR00368: H13-V35, V199-N208, F104-E129, N197-A211, V240-A247	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7511505CD1	443	S81 S173 T61 T351 T409 Y350 Y430	N220 N255	Ergosterol biosynthesis ERG4/ERG24 family: P7-F443	HMMER_PFAM
					Cytosolic domains: M1-A118, F169-H231, N323-M381; Transmembrane domains: W119-I141, I146-Y168, S232-W254, L300-A322, G382-M404; Non-cytosolic domains: F142-W145, N255-Q299, A405-F443	TMHMMER
					Ergosterol biosynthesis ERG4/ERG24 enzymes IPB001171: G181-R196, K199-V234, E256-P297, G315-R330, C348-C390, G391-F443	BLIMPS_BLOCKS
					REDUCTASE STEROL TRANSMEMBRANE OXIDOREDUCTASE BIOSYNTHESIS C14 C14REDUCTASE PROTEIN LAMIN B PD004179: L67-L396, A389-F443	BLAST_PRODOM
					REDUCTASE STEROL TRANSMEMBRANE OXIDOREDUCTASE BIOSYNTHESIS C14 LAMIN B RECEPTOR C14REDUCTASE PD149902: Y398-F443	BLAST_PRODOM
					DELTA7STEROL REDUCTASE OXIDOREDUCTASE PD140411: M1-A32, C34-T61	BLAST_PRODOM
					ERGOSTEROL BIOSYNTHESIS ERG4/ERG24 FAMILY DM01860|A53616|196-614: L67-F443	BLAST_DOMO
					ERGOSTEROL BIOSYNTHESIS ERG4/ERG24 FAMILY DM01860|P23913|190-607: K63-Y436	BLAST_DOMO
					ERGOSTEROL BIOSYNTHESIS ERG4/ERG24 FAMILY DM01860|JC4057|3-423: Y111-F443, A48-R330	BLAST_DOMO
					ERGOSTEROL BIOSYNTHESIS ERG4/ERG24 FAMILY DM01860|P38670|11-489: L67-G391, A389-F443	BLAST_DOMO

Table 3

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>11</td><td>7511552CD1</td><td>247</td><td>S34 S83 T45 T129 T155</td><td></td><td>Aldehyde dehydrogenase family IPB002086: P159-P172</td><td>BLIMPS_BLOCKS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE PD000218: E28-I216</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P00352|17-471: I18-A212</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P15437|17-471: Q19-A212</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P51977|17-471: Q19-A212</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P48644|17-471: Q19-A212</td><td>BLAST_DOMO</td></tr>
<tr><td>12</td><td>7511722CD1</td><td>237</td><td>S30 S214 Y161</td><td>N9 N186</td><td>Signal Peptide: M1-A26</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Signal Peptide: M1-G28</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN OF SECTION GLYOXYLATE-INDUCED GBD 5'REGION ORF4 PPHBRPOS INTERGENIC REGION PD150048: R5-G237</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>GLYOXYLATE; INDUCED; TRANSIENT; POTENTIAL; DM03647|I39562|1-257: R5-G237</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>GLYOXYLATE; INDUCED; TRANSIENT; POTENTIAL; DM03647|P30147|1-257: L4-G237</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>GLYOXYLATE; INDUCED; TRANSIENT; POTENTIAL; DM03647|C64108|1-257: R5-G237</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>GLYOXYLATE; INDUCED; TRANSIENT; POTENTIAL; DM03647|P36951|1-260: L4-G237</td><td>BLAST_DOMO</td></tr>
</table>

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7511489CD1	358	S39 S164 S188 S190 S297 T50 T81 T201 T306	N48	Squalene/phytoene synthase: L47-V275	HMMER_PFAM
					Squalene and phytoene synthases IPB002060: S51-T81, Y171-G182, M207-P232	BLIMPS_BLOCKS
					FARNESYLDIPHOSPHATE FARNESYL-TRANSFERASE SQUALENE SYNTHETASE SQS SS FPP: FPP MULTIFUNCTIONAL ENZYME TRANSFERASE PD023853:K276-H358	BLAST_PRODOM
					SYNTHASE TRANSFERASE BIOSYNTHESIS MULTIFUNCTIONAL ENZYME PHYTOENE SQUALENE CAROTENOID ISOPRENE PROTEIN PD001082: L47-E234, N246-V275	BLAST_PRODOM
					FARNESYLDIPHOSPHATE FARNESYLTRANSFERASE SQUALENE SYNTHETASE SQS SS FPP: FPP MULTIFUNCTIONAL ENZYME TRANSFERASE PD023850:M1-Y46	BLAST_PRODOM
					SQUALENE AND PHYTOENE SYNTHASES DM02785|P37268|1-403: M1-V235 P232-L345	BLAST_DOMO
					SQUALENE AND PHYTOENE SYNTHASES DM02785|P53800|1-408: L7-V235 P232-L337	BLAST_DOMO
					SQUALENE AND PHYTOENE SYNTHASES DM02785|P36596|1-435: M28-V235 P232-K299	BLAST_DOMO
					SQUALENE AND PHYTOENE SYNTHASES DM02785|P53799|1-409: L7-V235 P232-L342	BLAST_DOMO
					Squalene and phytoene synthases signature 1: Y171-L186	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Squalene and phytoene synthases signature 2: M207-P232	MOTIFS
14	7511497CD1	140	S102 S121 T132 Y48		DM08732 METHYLTRANSFERASE|P25627|10-189: P14-A89	BLAST_DOMO
					DM08732 METHYLTRANSFERASE|Q10162|10-189: P14-V97	BLAST_DOMO
15	7511498CD1	248	S194 Y48		ubiE/COQ5 methyltransferase family IPB000339: L58-S69, D82-R94	BLIMPS_BLOCKS
					METHYLTRANSFERASE TRANSFERASE SIMILAR S CEREVISIAE GENE YCR47C PD031003: L128-Q170	BLAST_PRODOM
					DM08732 METHYLTRANSFERASE|P25627|10-189: P14-S194	BLAST_DOMO
					DM08732 METHYLTRANSFERASE|Q10162|10-189: P14-S194	BLAST_DOMO
16	7511612CD1	153	S34 T59 T60		Signal_cleavage: M1-A35	SPSCAN
					Thiolases signatures: A104-A153	PROFILESCAN
					3KETOACYLCOA THIOLASE PEROXISOMAL PRECURSOR EC 2.3.1.16 BETA KETOTHIOLASE ACETYLCOA ACYL-TRANSFERASE 3OXOACYL COA FATTY ACID METABOLISM TRANSFERASE PEROXISOME TRANSIT PEPTIDE PD060746: M1-A36	BLAST_PRODOM
					TRANSFERASE SYNTHASE POLYKETIDE ACYLTRANSFERASE PROTEIN THIOLASE BIOSYNTHESIS ACETYLCOA ACID PUTATIVE PD000145: D37-A134	BLAST_PRODOM
					THIOLASES DM00548|P07871|34-421: S34-G135	BLAST_DOMO
					THIOLASES DM00548|S57792|48-430: D37-A134	BLAST_DOMO
					THIOLASES DM00548|S33637|45-434: D37-A134	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					THIOLASES DM00548|Q05493|26-413: D37-A134	BLAST_DOMO
17	7511624CD1	376	S6 S26 S57 S79 S123 T34 T55 T163 Y96 Y347		Signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A22	HMMER
					Signal Peptide: M1-A29	HMMER
					Signal Peptide: M1-A24	HMMER
					Aminotransferase class I and II: E111-T375	HMMER_PFAM
					Aminotransferases class-I IPB001511: G63-G72, L154-R168, S222-G235	BLIMPS_BLOCKS
					Aminotransferases class-I pyridoxal-phosphate attachment site: K202-Q253	PROFILESCAN
					Aspartate aminotransferase signature PR00799: L153-W172, L184-G198, C220-V240, S252-L277, I320-E338, I346-V364	BLIMPS_PRINTS
					AMINOTRANSFERASE PYRIDOXAL PHOSPHATE SYNTHASE TRANSFERASE ASPARTATE TRANSAMINASE 1AMINOCYCLOPROPANE1CARBOXYLATE A ACC PD001174: S30-L145	BLAST_PRODOM
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P12344|32-429: W32-K146, L121-K376	BLAST_DOMO
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P46643|32-429: W32-K146, P148-K376	BLAST_DOMO
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|S56678|29-427: W32-F144, P148-K376	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P46645|4-404: W33-L155, P148-K376	BLAST_DOMO
					Aminotransferases class-I pyridoxal-phosphate attachment site: S222-G235	MOTIFS
18	7511626CD1	387	S6 S26 S57 S79 T34 T55 T174 Y358		Signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A22	HMMER
					Signal Peptide: M1-A29	HMMER
					Signal Peptide: M1-A24	HMMER
					Aminotransferase class I and II: V76-T386	HMMER_PFAM
					Aminotransferases class-I IPB001511: G63-G72, L165-R179, S233-G246	BLIMPS_BLOCKS
					Aminotransferases class-I pyridoxal-phosphate attachment site: K213-Q264	PROFILESCAN
					Aspartate aminotransferase signature PR00799: L164-W183, L195-G209, C231-V251, S263-L288, I331-E349, I357-V375	BLIMPS_PRINTS
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P12344|32-429: P43-K387, W32-K82	BLAST_DOMO
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P46643|32-429: W32-K82, R81-K387	BLAST_DOMO
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|S56678|29-427: W32-K82, R81-K387	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					AMINOTRANSFERASES CLASS-I PYRIDOXAL-PHOSPHATE ATTACHMENT SI DM00997|P46248|52-452: V36-K82, G72-I382	BLAST_DOMO
					Aminotransferases class-I pyridoxal-phosphate attachment site: S233-G246	MOTIFS
19	7512885CD1	274	S146 S217 T6 T106 T150 T154 Y167		Signal_cleavage: M1-G65	SPSCAN
					Signal Peptide: M16-G33	HMMER
					Signal Peptide: M16-G36	HMMER
					Signal Peptide: M16-F38	HMMER
					Signal Peptide: M16-R39	HMMER
					Signal Peptide: M16-G46	HMMER
					short chain dehydrogenase: R51-R264	HMMER_PFAM
					Cytosolic domain: M1-T20 Transmembrane domain: S21-W43 Non-cytosolic domain: V44-S274	TMHMMER
					Alcohol dehydrogenase superfamily signature PR00080: G134-I145, G187-I195	BLIMPS_PRINTS
					Glucose/ribitol dehydrogenase family signature PR00081: V54-V71, G134-I145, M181-G197	BLIMPS_PRINTS
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P54554|1-237: I57-A213, P218-L242	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P00335|9-246: L50-I201, A206-L242	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|Q10783|48-289: L50-A213	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P14802|1-235: L50-S244	BLAST_DOMO
						MOTIFS
20	7511965CD1	293	S8 S103 S133 S204 S259 S282 T3 T20 T82 T96 T118 T137 T222 Y40 Y47	N234	Signal_cleavage: M1-G26	SPSCAN
					Aldo/keto reductase family: L7-W273	HMMER_PFAM
					Ribosomal protein L10E IPB001197: I182-N234	BLIMPS_BLOCKS
					Aldo/keto reductase family IPB001395: L7-L18, A35-Q60, V68-W80, T96-L109, K154-S188, I214-M263	BLIMPS_BLOCKS
					Aldo/keto reductase family signatures: I211-A275	PROFILESCAN
					Aldo-keto reductase signature PR00069: A35-I59, K95-P113, V158-Y187, L205-F229	BLIMPS_PRINTS
					PROTEIN OXIDOREDUCTASE REDUCTASE CHANNEL NADP DEHYDROGENASE ALDEHYDE IONIC ALDOSE PUTATIVE PD000288: V28-L153, E143-N272, L7-D72, E249-M263	BLAST_PRODOM
					ALDO/KETO REDUCTASE FAMILY DM00192|P07943|1-297: A2-N160, E143-R274	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|I49484|2-298: A2-N160, E143-R274	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|P31210|7-308: V5-N160, E143-R274	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|P51857|7-308: V5-N160, E143-R274	BLAST_DOMO
					Aldo/keto reductase family signature 1: G39-G56	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7512403CD1	376	S14 S184 S320 S337 T80 T90 T143 T210 T275		Aldehyde dehydrogenase family: M1-S320	HMMER_PFAM
					Cytosolic domain: E374-Y376 Transmembrane domain: L355-A373 Non-cytosolic domain: M1-K354	TMHMMER
					Gamma-glutamyl phosphate reductase IPB000965: G15-E67, L106-T143, E222-L250	BLIMPS_BLOCKS
					Aldehyde dehydrogenase family IPB002086: P101-P114, T201-L250, V279-G303	BLIMPS_BLOCKS
					ALDEHYDE DEHYDROGENASE LIKE PROTEIN PD065537: L139-I365	BLAST_PRODOM
					DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE PD000218: V5-Q153, G161-K319	BLAST_PRODOM
					DEHYDROGENASE FATTY ALDEHYDE MICROSOMAL CLASS OXIDOREDUCTASE NAD ENDOPLASMIC RETICULUM MICROSOME PD020122: L321-K372	BLAST_PRODOM
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P51648|1-405: M1-Y160, Q157-G297	BLAST_DOMO
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P30838|1-408: V5-Y160, Q157-G297	BLAST_DOMO
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P43353|1-408: V5-Y160, K149-G297	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P12693|1-432: L3-L151, L139-G297	BLAST_DOMO
22	7512564CD1	248	S106 T57	N228	Signal_cleavage: M1-A20	SPSCAN
					Signal Peptide: M5-P22	HMMER
					Signal Peptide: M5-K26	HMMER
					Signal Peptide: M1-R25	HMMER
					Signal Peptide: M1-K26	HMMER
					short chain dehydrogenase: P40-R233	HMMER_PFAM
					Alcohol dehydrogenase superfamily signature PR00080: K119-V130	BLIMPS_PRINTS
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S42651|28-318: V45-T157, G152-L243	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S30167|82-368: V44-G152, H150-I187	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S39394|69-356: V43-G152, H150-I187	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P13653|69-356: V43-G152, H150-I187	BLAST_DOMO
23	7512646CD1	198	S86 S116 T146 T195		Signal_cleavage: M1-A51	SPSCAN
					AhpC/TSA family: V65-R198	HMMER_PFAM
					PROTEIN ANTIOXIDANT PEROXIDASE REDUCTASE PD00210: V96-E111	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN ANTIOXIDANT PEROXIDASE REDUCTASE PROBABLE OXIDOREDUCTASE THIOREDOXIN PEROXIDE THIOREDOXIN-DEPENDENT ALKYL PD000498: A64-A182	BLAST_PRODOM
					MITOCHONDRIAL THIOREDOXIN-DEPENDENT PEROXIDE REDUCTASE PRECURSOR ANTIOXIDANT PROTEIN AOP1 MITOCHONDRION TRANSIT PD019666: M1-A64	BLAST_PRODOM
					ALKYL HYDROPEROXIDASE C22 PROTEIN DM00437|P20108|62-250: H61-R184	BLAST_DOMO
					ALKYL HYDROPEROXIDASE C22 PROTEIN DM00437|JC2258|1-188: A62-R184	BLAST_DOMO
					ALKYL HYDROPEROXIDASE C22 PROTEIN DM00437|A57716|4-192: A69-R184	BLAST_DOMO
					ALKYL HYDROPEROXIDASE C22 PROTEIN DM00437|P35703|4-193: A69-R184	BLAST_DOMO
						MOTIFS
24	7512700CD1	270	S7 S234 T25 T49 T168 T236 Y76	N176 N196	Signal_cleavage: M1-G24	SPSCAN
					Aldo/keto reductase family: L12-V250	HMMER_PFAM
					Aldo/keto reductase family IPB001395: L12-L23, T49-I62, T94-P127, K130-S164, L191-M240	BLIMPS_BLOCKS
					Aldo/keto reductase family signatures: D73-Q137, L188-R251	PROFILESCAN
					Aldo-keto reductase signature PR00069: R48-P66, M98-F115, V134-Y163, L182-F206	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN OXIDOREDUCTASE REDUCTASE CHANNEL NADP DEHYDROGENASE ALDEHYDE IONIC ALDOSE PUTATIVE PD000288: L12-Y26, T25-L255	BLAST_PRODOM
					ALDO/KETO REDUCTASE FAMILY DM00192|P51857|7-308: S7-S31, L32-V253	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|P31210|7-308: S7-K30, L32-V253	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|JH0575|5-305: S7-M41, L32-F252	BLAST_DOMO
					ALDO/KETO REDUCTASE FAMILY DM00192|P51652|5-305: R9-L32, L32-F252	BLAST_DOMO
					Aldo/keto reductase family signature 2: M98-F115	MOTIFS
					Aldo/keto reductase family putative active site signature: I215-I230	MOTIFS
25	7512707CD1	306	S23 S28 S84 S160 S283 S295 T34 T192 T273 Y141	N222	Signal_cleavage: M1-P16	SPSCAN
					GMC oxidoreductase IPB000172: D37-G55	BLIMPS_BLOCKS
					UbiH/COQ6 monooxygenase family IPB000689: D37-L54, K202-L236	BLIMPS_BLOCKS
					Aromatic-ring hydroxylase (flavoprotein monooxygenase) signature PR00420: D37-H59, Q200-R215	BLIMPS_PRINTS
					PROTEIN OXIDOREDUCTASE MONOOXYGENASE FLAVOPROTEIN FAD UBIQUINONE BIOSYNTHESIS PUTATIVE MONOOXGENASE COQ6 PD009774: D37-E157	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					UBIQUINONE BIOSYNTHESIS MONOOXGENASE COQ6 OXIDOREDUCTASE MONOOXYGENASE FLAVOPROTEIN FAD PUTATIVE PROTEIN PD155418: I205-L281	BLAST_PRODOM
26	7512710CD1	423	S23 S28 S115 S238 S254 S282 T34 T147 T228 T365 T369 Y96	N177	Signal_cleavage: M1-P16	SPSCAN
					Monooxygenase: Q155-D383	HMMER_PFAM
					Cytosolic domain: M1-D37 Transmembrane domain: V38-D60 Non-cytosolic domain: H61-K423	TMHMMER
					UbiH/COQ6 monooxygenase family IPB000689: D37-L54, K157-L191, F301-L344	BLIMPS_BLOCKS
					Glucose-methanol-choline oxidoreductase IPB000172: D37-G55	BLIMPS_BLOCKS
					Fumarate reductase/succinate dehydrogenase, FAD-binding site IPB000464: D37-A58	BLIMPS_BLOCKS
					FAD-dependent glycerol-3-phosphate dehydrogenase family signature PR01001: Y36-A48	BLIMPS_PRINTS
					Aromatic-ring hydroxylase (flavoprotein monooxygenase) signature PR00420: D37-W59, Q155-R170, R313-P328, P328-L344, Y367-D383	BLIMPS_PRINTS
					Flavin-containing amine oxidase signature PR00757: D37-F56	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					UBIQUINONE BIOSYNTHESIS MONOOXGENASE COQ6 OXIDOREDUCTASE MONOOXYGENASE FLAVOPROTEIN FAD PUTATIVE PROTEIN PD155418: I160-L236	BLAST_PRODOM
					FAD NP_BIND DM00811|P25535|207-346: G210-F252, V290-L379	BLAST_DOMO
					FAD NP_BIND DM00811|P53318|282-431: P287-L379	BLAST_DOMO
					FAD NP_BIND DM00811|P25534|205-343: A295-L379	BLAST_DOMO
					ubiH/COQ6 monooxygenase family signature: H327-D340	MOTIFS
27	7512884CD1	88	T11		DIOXYGENASE CYSTEINE CDO OXIDOREDUCTASE IRON CDOI SIMILAR HUMAN PD016803: L15-G82	BLAST_PRODOM
					Riboflavin synthase alpha chain family signature: L19-E31	MOTIFS
28	7512931CD1	458	S36 S45 S160 S209 S252 S256 S318 S413 S443 T174 T226		Signal Peptide: M1-A24	HMMER
					Flavin containing amine oxidoreductase: I12-L452	HMMER_PFAM
					proto_IX_ox: PROTO-PORPHYRINOGEN oxidase: M1-S458	HMMER_TIGRFAM
					D-amino acid oxidase, DAO IPB000927: T4-A16	BLIMPS_BLOCKS
					Glucose-methanol-choline oxidoreductase IPB000172: T4-S22	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					OXIDASE PROTO-PORPHYRINOGEN OXIDOREDUCTASE PPO PORPHYRIN BIOSYNTHESIS FLAVOPROTEIN FAD HEME MITOCHONDRION PD004706: V101-R188, W208-S353, I53-K117	BLAST_PRODOM
					OXIDASE PROTO-PORPHYRINOGEN OXIDOREDUCTASE PPO PORPHYRIN BIOSYNTHESIS FLAVOPROTEIN FAD HEME MITOCHONDRION PD006423: G383-G453	BLAST_PRODOM
					PROTO-PORPHYRINOGEN OXIDASE PPO PORPHYRIN BIOSYNTHESIS HEME OXIDOREDUCTASE FLAVOPROTEIN FAD MITOCHONDRION PD041529: W354-H396	BLAST_PRODOM
					PROTO-PORPHYRINOGEN; OXIDASE; DM03709|P50336|3-340: R3-P118, G111-D321	BLAST_DOMO
					PROTO-PORPHYRINOGEN; OXIDASE; DM04318|P50336|365-476: V346-S458	BLAST_DOMO
					PROTO-PORPHYRINOGEN; OXIDASE; DM03709|Q10062|1-353: T4-L112, A104-S318	BLAST_DOMO
					PROTO-PORPHYRINOGEN; OXIDASE; DM03709|P32397|6-344: R47-D321	BLAST_DOMO
					Cell attachment sequence: R88-D90	MOTIFS
29	7512933CD1	185	S36 S45 S140 S170		Signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-A24	HMMER
					Glucose-methanol-choline oxidoreductase IPB000172: T4-S22	BLIMPS_BLOCKS
					Flavin-containing amine oxidase signature PR00757: T4-R23	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Flavin-containing monooxygenase (FMO) signature PR00370: R3-Y19	BLIMPS_PRINTS
					OXIDASE PROTO-PORPHYRINOGEN OXIDOREDUCTASE PPO PORPHYRIN BIOSYNTHESIS FLAVOPROTEIN FAD HEME MITOCHONDRION PD006423: G110-G180	BLAST_PRODOM
					PROTO-PORPHYRINOGEN OXIDASE PPO PORPHYRIN BIOSYNTHESIS HEME OXIDOREDUCTASE FLAVOPROTEIN FAD MITOCHONDRION PD041529: W81-H123	BLAST_PRODOM
					PROTO-PORPHYRINOGEN; OXIDASE; DM04318|P50336|365-476: V75-S185	BLAST_DOMO
					PROTO-PORPHYRINOGEN; OXIDASE; DM03709|P50336|3-340: R3-V75	BLAST_DOMO
30	7512942CD1	140	S95 S125		Signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-A24	HMMER
					D-amino acid oxidase, DAO IPB000927: T4-A16	BLIMPS_BLOCKS
					Glucose-methanol-choline oxidoreductase IPB000172: T4-S22	BLIMPS_BLOCKS
					Flavin-containing amine oxidase signature PR00757: T4-R23	BLIMPS_PRINTS
					Flavin-containing monooxygenase (FMO) signature PR00370: R3-Y19	BLIMPS_PRINTS
					OXIDASE PROTO-PORPHYRINOGEN OXIDOREDUCTASE PPO PORPHYRIN BIOSYNTHESIS FLAVOPROTEIN FAD HEME MITOCHONDRION PD006423: G65-G135	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTO-PORPHYRINOGEN OXIDASE PPO PORPHYRIN BIOSYNTHESIS HEME OXIDOREDUCTASE FLAVOPROTEIN FAD MITOCHONDRION PD041529: W36-H78	BLAST_PRODOM
					PROTO-PORPHYRINOGEN; OXIDASE; DM04318|P50336|365-476: V30-S140	BLAST_DOMO
					PROTO-PORPHYRINOGEN; OXIDASE; DM03709|P50336|3-340: R3-M69	BLAST_DOMO
31	7513736CD1	281	S20 S56 S112 S184 S267 T34 T92 T178 T212	N176	Signal_cleavage: M1-S18	SPSCAN
					FMN-dependent dehydrogenase: A13-F278	HMMER_PFAM
					FMN-dependent alpha-hydroxy acid dehydrogenases IPB000262: V204-E247	BLIMPS_BLOCKS
					OXIDASE OXIDOREDUCTASE DEHYDROGENASE FLAVOPROTEIN FMN LLACTATE GLYCOLATE PROTEIN CYTOCHROME CHAIN PD003201: A13-S184	BLAST_PRODOM
					OXIDASE OXIDOREDUCTASE PROTEIN FLAVOPROTEIN FMN DEHYDROGENASE CYTOCHROME LLACTATE CHAIN GLYCOLATE PD006736: G208-N272	BLAST_PRODOM
					FMN-DEPENDENT ALPHA-HYDROXY ACID DEHYDROGENASES DM01725|P00175|201-561: K199-L229, S2-R170, V198-S279	BLAST_DOMO
					FMN-DEPENDENT ALPHA-HYDROXY ACID DEHYDROGENASES DM01725|P05414|2-357: L6-Q183, P185-N274	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					FMN-DEPENDENT ALPHA-HYDROXY ACID DEHYDROGENASES DM01725|P09437|186-540: E114-N138, L3-R170, I188-V276	BLAST_DOMO
					FMN-DEPENDENT ALPHA-HYDROXY ACID DEHYDROGENASES DM01725|Q07523|1-350: L3-K187, I188-R280	BLAST_DOMO
					Bacterial regulatory proteins, lysR family signature: R94-L124	MOTIFS
32	7512779CD1	463	S40 S62 S66 S81 S95 S142 S207 T186 T285	N112	Pyridoxal-dependent decarboxylase conserved domain: P49-P417	HMMER_PFAM
					Pyridoxal-dependent decarboxylase family IPB002129: G144-Q166, P181-L203, W269-G278, G292-D320	BLIMPS_BLOCKS
					GLUTAMATE DECARBOXYLASE PD114206: L10-N112	BLAST_PRODOM
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE MULTIGENE FAMILY DOPA GLUTAMATE ACID AROMATICL-AMINOACID PD001960: P49-E406	BLAST_PRODOM
					DECARBOXYLASE CYSTEINE SULFINOALANINE CYSTEINESULFINATE LYASE PYRIDOXAL PHOSPHATE SULFINIC ACID SULFINATE PD153956: M1-E48	BLAST_PRODOM
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568|JH0827|84-575: L10-E406	BLAST_DOMO
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568|P14748|101-591: E5-E406	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568|Q05329|95-582: S4-E406	BLAST_DOMO
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568|S55689|5-478: L7-P407	BLAST_DOMO
					Eukaryotic putative RNA-binding region RNP-1 signature: K196-L203	MOTIFS
33	7512877CD1	149	S70 S77 S145 T10 T106		3-hydroxyacyl-CoA dehydrogenase, NAD binding: E43-Q149	HMMER_PFAM
					3-hydroxyacyl-CoA dehydrogenase IPB002135: D118-Q150	BLIMPS_BLOCKS
					DEHYDROGENASE PROTEIN 3HYDROXY-ACYLCOA FATTY ACID METABOLISM OXIDOREDUCTASE 3HYDROXY-BUTYRYLCOA INCLUDES: ENOYLCOA PD001126: V46-K140	BLAST_PRODOM
					3-HYDROXYACYL-COA DEHYDROGENASE DM00842|P00348|11-306: V46-E144	BLAST_DOMO
					3-HYDROXYACYL-COA DEHYDROGENASE DM00842|P41938|18-308: V46-D139	BLAST_DOMO
34	7512782CD1	203	S110 S188		Signal_cleavage: M1-A44	SPSCAN
					Signal Peptide: M1-G29	HMMER
					Quinolinate phosphoribosyl transferase, C-terminal domain: A62-K190	HMMER_PFAM
					Quinolinate phosphoribosyl transferase, N-terminal domain: C27-S110	HMMER_PFAM
					Nicotinate-nucleotide pyrophosphoryl a (nadC): L18-L191	HMMER_TIGRFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					TRANSFERASE NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE QUINOLINATE CARBOXYLATING PHOSPHORIBOSYL-TRANSFERASE DECARBOXYLATING QAPRTASE GLYCOSYL-TRANSFERASE PROBABLE PD003988: W22-R95, V87-P147	BLAST_PRODOM
					PYROPHOSPHORYLASE; NICOTINATE; NUCLEOTIDE; TRANSPORT; DM02806|P43619|1-274: V19-P83, V87-V172	BLAST_DOMO
35	7512784CD1	166	S73 S151		Signal_cleavage: M1-A44	SPSCAN
					Signal Peptide: M1-G29	HMMER
					Quinolinate phosphoribosyl transferase, C-te: T14-K153	HMMER_PFAM
					Nicotinate-nucleotide pyrophosphoryl a (nadC): L18-L154	HMMER_TIGRFAM
					TRANSFERASE NICOTINATENUCLEOTIDE PYROPHOSPHORYLASE QUINOLINATE CARBOXYLATING PHOSPHORIBOSYL-TRANSFERASE DECARBOXYLATING QAPRTASE GLYCOSYLTRANSFERASE PROBABLE PD003988: A53-P110	BLAST_PRODOM
					PYROPHOSPHORYLASE; NICOTINATE; NUCLEOTIDE; TRANSPORT; DM02806|P43619|1-274: V19-G52, L46-V135	BLAST_DOMO
36	7512794CD1	48	T10		S-adenosylmethionine decarboxylase IPB001985: F7-F18	BLIMPS_BLOCKS
					DECARBOXYLASE BIOSYNTHESIS S-ADENOSYLMETHIONINE PD02625: F7-F18	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					S-ADENOSYLMETHIONINE DECARBOXYLASE BETA CHAIN DM01913|I48237|1-333: M1-L42	BLAST_DOMO
37	7512886CD1	244	S25 S172 T87 T207		Signal_cleavage: M1-G17	SPSCAN
					Signal Peptide: M1-G17	HMMER
					Signal Peptide: M1-L19	HMMER
					Signal Peptide: M1-W20	HMMER
					Signal Peptide: M1-D23	HMMER
					Signal Peptide: M4-D23	HMMER
					Cytosolic domain: M1-M4 Transmembrane domain: W5-W22 Non-cytosolic domain: D23-P244	TMHMMER
					Sulfatase IPB000917: M55-R66	BLIMPS_BLOCKS
					PIGL PROTEIN CAT8ATP13 INTERGENIC REGION URF2 PD039078: L45-R199	BLAST_PRODOM
38	7512929CD1	115	S42 S53 S84 T74		Signal_cleavage: M1-G35	SPSCAN
					Signal Peptide: M18-S32	HMMER
					Signal Peptide: M18-G35	HMMER
					Signal Peptide: R16-G35	HMMER
					Signal Peptide: N14-G35	HMMER
					Signal Peptide: M1-G35	HMMER
					Cytosolic domain: M1-K19 Transmembrane domain: L20-V38 Non-cytosolic domain: N39-H115	TMHMMER
					Tau and MAP proteins tubulin-binding domain signature: V10-I66	PROFILESCAN
					Arginine ADP-ribosyltransferase signature PR00970: M18-S32	BLIMPS_PRINTS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
39/7511289CB1 1798	1-105, 1-126, 1-210, 1-471, 1-644, 1-1792, 15-219, 18-219, 18-297, 24-579, 54-320, 105-544, 130-714, 152-353, 168-438, 174-433, 175-436, 179-559, 182-615, 196-708, 199-461, 202-697, 208-771, 225-721, 227-711, 227-713, 227-714, 228-523, 236-714, 244-444, 244-451, 247-770, 257-729, 266-714, 279-577, 305-614, 305-713, 315-621, 346-807, 346-865, 403-694, 434-530, 453-715, 476-715, 490-694, 518-715, 533-715, 713-963, 722-1154, 723-1154, 735-1154, 767-1038, 767-1155, 770-1223, 771-1297, 791-1068, 803-1056, 812-1241, 822-1088, 822-1292, 842-1154, 845-1489, 860-1311, 879-1122, 881-1470, 886-1148, 912-1187, 949-1280, 952-1096, 989-1231, 1010-1620, 1029-1225, 1031-1778, 1036-1624, 1042-1274, 1079-1349, 1080-1638, 1097-1414, 1126-1400, 1145-1798, 1162-1421, 1163-1389, 1167-1490, 1168-1786, 1191-1446, 1204-1786, 1205-1798, 1213-1438, 1213-1485, 1243-1429, 1274-1467, 1277-1668, 1277-1798, 1279-1735, 1281-1782, 1282-1786, 1283-1770, 1283-1790, 1296-1669, 1296-1782, 1298-1732, 1298-1761, 1311-1786, 1318-1786, 1325-1732,
	1325-1786, 1326-1596, 1326-1732, 1330-1732, 1331-1732, 1334-1732, 1335-1798, 1336-1784, 1337-1732, 1337-1786, 1339-1732, 1343-1761, 1345-1732, 1346-1798, 1348-1732, 1352-1798, 1353-1732, 1354-1786, 1358-1786, 1360-1786, 1366-1798, 1371-1732, 1373-1582, 1373-1732, 1377-1778, 1377-1786, 1379-1732, 1385-1776, 1387-1778, 1388-1798, 1389-1732, 1399-1732, 1402-1732, 1408-1797, 1408-1798, 1409-1786, 1411-1732, 1414-1732, 1421-1794, 1424-1732, 1429-1795, 1435-1778, 1441-1722, 1441-1732, 1456-1786, 1457-1667, 1461-1718, 1461-1732, 1461-1737, 1461-1784, 1461-1785, 1461-1786, 1462-1737, 1462-1740, 1462-1786, 1466-1798, 1467-1762, 1483-1646, 1486-1798, 1487-1732, 1490-1732, 1490-1796, 1490-1798, 1492-1732, 1494-1732, 1500-1736, 1500-1794, 1502-1732, 1512-1732, 1522-1732, 1543-1732, 1544-1731, 1544-1732, 1553-1720, 1554-1798, 1561-1732, 1563-1732, 1576-1732, 1576-1789, 1576-1798, 1598-1778, 1604-1714, 1623-1795, 1636-1732, 1640-1798, 1644-1786, 1655-1778, 1655-1780
40/7511056CB1/ 929	1-271, 1-432, 1-569, 4-499, 70-563, 81-525, 351-867, 570-929
41/7511567CB1/ 2163	1-2108, 39-857, 64-416, 64-523, 65-599, 67-423, 67-495, 67-499, 69-660, 71-331, 71-605, 71-636, 71-686, 74-629, 81-482, 93-313, 93-377, 93-437, 93-496, 93-499, 93-547, 93-548, 93-554, 93-674, 95-544, 154-421, 186-478, 194-402, 194-472, 210-459, 210-471, 210-656, 226-821, 227-726, 292-454, 292-497, 297-621, 317-758, 338-1015, 352-742, 352-804, 352-816, 375-633, 388-1015, 395-659, 402-659, 402-722, 415-959, 440-1190, 451-669, 455-1017, 466-818, 466-889, 467-1039, 473-738, 487-798, 489-806, 514-892, 535-892, 546-897, 546-1143, 552-791, 565-1132, 587-838, 587-846, 596-735, 606-882, 622-923, 645-1332, 673-868, 673-1055, 690-1209, 694-923, 700-930, 712-940, 712-1282, 712-1378, 722-989, 722-1192, 724-1121, 724-1331, 762-879, 762-917, 762-1129, 762-1204, 762-1246, 762-1316, 762-1362, 762-1416, 763-1132, 771-1373, 781-1019, 785-1354, 787-1144, 813-1365, 841-1063, 848-1146, 853-1279, 876-1435, 877-1132, 881-1148, 886-1122, 886-1183, 887-1187, 888-1137, 888-1179, 888-1539, 898-1132, 911-1122, 919-1140, 920-1142,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	923-1162, 923-1397, 929-1162, 936-1377, 940-1142, 947-1159, 965-1374, 967-1227, 967-1406, 974-1533, 978-1249, 982-1565, 983-1490, 991-1490, 1028-1587, 1029-1566, 1036-1301, 1038-1446, 1040-1176, 1041-1406, 1069-1490, 1072-1571, 1088-1497, 1098-1567, 1103-1490, 1105-1630, 1111-1410, 1123-1616, 1124-1389, 1129-1412, 1132-1403, 1137-1418, 1142-1353, 1142-1410, 1142-1423, 1145-1580, 1146-1404, 1156-1594, 1179-1666, 1185-1414, 1185-1420, 1185-1472, 1195-1433, 1196-1436, 1202-1565, 1208-1502, 1221-1458, 1223-1522, 1228-1571, 1228-1611, 1257-1535, 1271-1391, 1272-1552, 1296-1628, 1307-1579, 1307-1588, 1316-1522, 1330-1628, 1362-1593, 1420-1666, 1425-1655, 1433-1652, 1473-1706, 1664-1758, 1664-1810, 1664-1874, 1664-2031, 1664-2066, 1664-2090, 1664-2103, 1664-2106, 1664-2107, 1664-2125, 1664-2129, 1665-2139, 1666-2117, 1667-2069, 1667-2139, 1668-2111, 1669-2107, 1670-2079, 1670-2131, 1680-1982, 1680-2075, 1680-2100, 1684-1909, 1685-1998, 1685-2133, 1686-1812, 1687-2075, 1687-2126, 1688-2085, 1691-2090, 1691-2122,
	1692-2099, 1692-2114, 1697-1982, 1697-2077, 1697-2122, 1698-1956, 1701-2107, 1709-1982, 1716-1954, 1719-2107, 1723-1929, 1723-1997, 1723-2049, 1723-2111, 1723-2125, 1736-2122, 1742-1919, 1742-1928, 1742-2007, 1744-2108, 1746-2067, 1747-1999, 1747-2107, 1753-2107, 1756-1918, 1762-2125, 1773-2128, 1776-2106, 1777-1973, 1778-2107, 1786-2108, 1812-2101, 1812-2138, 1835-2107, 1839-2108, 1846-2104, 1852-2107, 1852-2108, 1852-2111, 1856-2102, 1857-2102, 1858-1978, 1859-2130, 1860-2113, 1865-2107, 1868-2107, 1876-2133, 1883-2102, 1883-2103, 1883-2105, 1883-2107, 1883-2108, 1902-2128, 1904-2138, 1905-2108, 1916-2107, 1916-2113, 1925-2107, 1930-2127, 1953-2163, 1961-2108, 1964-2085, 1975-2107, 1984-2102, 1990-2112
42/7511651CB1/ 1260	1-153, 1-248, 13-261, 13-1252, 16-315, 16-318, 18-272, 19-262, 20-308, 24-306, 43-286, 47-268, 47-274, 59-323, 61-252, 63-266, 64-281, 64-323, 74-323, 78-297, 289-728, 323-582, 328-870, 340-569, 353-594, 354-724, 360-691, 360-830, 389-661, 395-599, 410-683, 424-669, 433-675, 453-690, 458-1108, 473-716, 483-744, 483-745, 504-1181, 508-768, 516-1252, 523-784, 523-792, 523-814, 533-804, 561-799, 561-803, 565-866, 583-770, 606-873, 608-820, 616-1252, 625-852, 626-1082, 634-822, 637-1134, 647-894, 647-942, 653-1249, 654-1144, 658-1252, 668-1151, 674-939, 682-1023, 689-993, 689-1252, 697-1172, 708-958, 717-1252, 719-1095, 727-1011, 744-1252, 747-983, 748-1223, 750-1217, 777-1053, 777-1252, 782-975, 791-1043, 793-1083, 794-1252, 800-1252, 801-1040, 801-1252, 805-1252, 808-1252, 816-1252, 824-1252, 828-1252, 835-1076, 835-1252, 839-1252, 840-1252, 843-1252, 844-1252, 848-1114, 851-1252, 857-1252, 858-1252, 860-1252, 867-1252, 874-1252, 876-1252, 877-1252, 878-1155, 878-1252, 879-1252, 880-1252, 882-1252, 893-1252,
	895-1190, 907-1252, 910-1135, 915-1252, 920-1252, 930-1252, 933-1218, 937-1198, 943-1207, 953-1246, 954-1252, 955-1252, 957-1220, 958-1223, 964-1156, 992-1252, 995-1252, 998-1252, 999-1224, 1001-1218, 1038-1252, 1040-1252, 1048-1252, 1060-1252, 1062-1252, 1069-1252, 1086-1196, 1086-1201, 1086-1252, 1086-1260, 1087-1252, 1088-1254, 1088-1260, 1089-1252, 1099-1252, 1107-1260, 1117-1252, 1122-1252, 1182-1252

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
43/7511881CB1/ 772	1-772, 104-706
44/7512181CB1/ 911	1-222, 1-273, 1-597, 4-671, 9-123, 9-256, 9-258, 9-332, 9-541, 9-692, 9-902, 12-554, 12-572, 15-177, 15-366, 18-712, 18-773, 19-712, 20-609, 21-246, 21-275, 23-223, 23-303, 24-656, 26-406, 26-533, 27-371, 34-299, 122-877, 167-545, 179-579, 183-417, 183-455, 216-679, 222-719, 223-427, 236-467, 267-524, 419-719, 456-719, 712-911, 804-888
45/7511726CB1/ 1170	1-464, 1-543, 8-278, 11-224, 12-324, 15-268, 21-189, 21-260, 21-277, 21-288, 21-364, 21-546, 21-565, 23-294, 23-316, 23-471, 24-272, 24-305, 24-536, 24-624, 24-646, 25-251, 25-298, 27-256, 27-314, 27-319, 30-172, 30-178, 30-218, 30-230, 30-247, 30-250, 30-264, 30-271, 30-272, 30-277, 30-284, 30-286, 30-289, 30-290, 30-293, 30-296, 30-300, 30-302, 30-305, 30-316, 30-318, 30-319, 30-320, 30-322, 30-330, 30-342, 30-346, 30-474, 30-501, 30-565, 30-577, 30-613, 30-614, 30-623, 30-632, 30-650, 30-660, 31-164, 31-262, 31-1170, 32-290, 33-131, 33-155, 33-179, 33-181, 33-236, 33-238, 33-253, 33-265, 33-269, 33-275, 33-276, 33-283, 33-285, 33-287, 33-295, 33-315, 33-387, 33-425, 33-452, 33-492, 33-494, 33-496, 33-556, 33-560, 33-580, 33-609, 33-635, 33-637, 33-641, 33-660, 33-664, 34-200, 34-274, 34-312, 34-340, 34-441, 34-606, 34-660, 35-288, 35-295, 35-311, 35-537, 35-609, 36-181, 36-269, 37-208, 37-269, 37-281, 37-286, 37-289, 37-324, 37-332, 37-343, 37-351, 37-442, 37-446, 37-549, 37-565, 37-571, 37-581, 37-588,
	37-603, 37-604, 39-219, 39-259, 39-277, 39-279, 39-306, 39-307, 39-311, 39-316, 39-321, 39-326, 39-327, 39-339, 39-362, 39-514, 39-620, 39-621, 39-638, 40-289, 40-301, 40-440, 41-269, 41-295, 41-618, 41-641, 41-649, 42-312, 42-316, 42-330, 43-209, 43-244, 43-290, 43-295, 43-299, 43-340, 44-112, 44-252, 44-261, 44-263, 44-268, 44-269, 44-276, 44-278, 44-279, 44-284, 44-290, 44-291, 44-292, 44-293, 44-297, 44-302, 44-305, 44-306, 44-315, 44-330, 44-337, 44-338, 44-347, 44-348, 44-355, 44-366, 44-486, 44-499, 44-558, 44-564, 44-628, 44-632, 44-641, 44-659, 44-660, 45-219, 45-239, 45-287, 45-297, 45-313, 45-315, 45-323, 45-324, 45-341, 45-351, 45-429, 46-282, 46-329, 46-337, 46-355, 47-209, 47-243, 47-250, 47-302, 47-322, 47-610, 50-290, 50-314, 51-235, 51-470, 52-241, 52-337, 53-315, 53-317, 53-341, 53-382, 53-587, 54-620, 55-291, 55-637, 57-341, 57-348, 57-641, 59-282, 60-162, 62-299, 62-315, 63-331, 63-334, 65-328, 65-336, 65-493, 66-648, 69-311, 69-602, 70-367, 71-538, 72-328, 72-643, 74-333, 74-334, 74-351,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	74-377, 75-269, 75-319, 75-619, 76-330, 76-365, 76-643, 76-660, 88-327, 91-438, 93-354, 117-508, 120-660, 121-540, 147-424, 153-660, 156-332, 156-427, 159-384, 166-434, 166-630, 167-354, 168-260, 178-313, 184-398, 184-405, 184-455, 185-438, 190-495, 193-483, 193-492, 197-316, 200-479, 201-386, 204-513, 206-465, 206-476, 209-660, 212-549, 218-464, 232-402, 232-457, 233-395, 233-494, 233-514, 236-495, 242-508, 242-535, 245-459, 245-523, 255-462, 261-520, 261-538, 272-471, 275-471, 278-497, 317-511, 317-580, 317-595, 317-600, 327-590, 327-631, 329-537, 329-660, 334-502, 334-542, 334-587, 341-521, 344-629, 344-657, 349-660, 350-487, 352-653, 357-596, 358-576, 358-593, 361-573, 362-627, 366-598, 368-628, 369-660, 373-631, 373-645, 377-597, 377-606, 378-660, 380-545, 382-660, 386-579, 398-632, 400-628, 407-657, 412-672, 412-1078, 416-660, 417-653, 444-660, 446-660, 476-555, 494-600, 655-1170, 657-818, 657-871, 657-882, 657-886, 657-889, 657-906, 657-913, 657-931, 657-1067, 657-1068, 657-1115, 657-1138, 657-1170,
	658-843, 659-901, 659-911, 660-922, 660-1170, 661-1170, 662-892, 663-1170, 667-1170, 673-1170, 674-945, 674-953, 675-1028, 678-788, 678-894, 681-903, 681-966, 681-1169, 682-995, 682-1082, 682-1161, 683-932, 684-1021, 688-900, 688-981, 688-1087, 688-1092, 688-1125, 688-1167, 690-951, 690-1064, 690-1114, 691-902, 691-903, 691-1151, 692-947, 692-1009, 692-1075, 693-1134, 694-937, 694-950, 695-922, 695-952, 697-989, 698-940, 698-1129, 698-1170, 699-893, 699-1014, 700-904, 700-956, 700-1076, 700-1170, 701-916, 701-941, 702-1015, 702-1102, 702-1165, 702-1170, 705-1170, 707-981, 707-1164, 712-997, 712-1170, 713-1170, 714-1170, 715-1170, 717-1164, 718-880, 718-1004, 718-1170, 719-940, 719-982, 719-999, 719-1006, 719-1008, 720-1170, 721-1014, 721-1170, 723-1170, 724-1170, 725-1170, 726-1170, 727-1170, 728-956, 728-1170, 729-1016, 729-1136, 729-1170, 730-965, 730-1170, 731-1016, 731-1170, 732-1164, 734-972, 734-983, 734-1010, 734-1026, 734-1027, 734-1040, 734-1106, 734-1130, 734-1136, 734-1170, 735-1170, 736-1170,
	737-893, 737-998, 738-1003, 738-1170, 739-1170, 742-1170, 744-979, 744-983, 744-984, 744-1170, 745-985, 745-987, 745-1170, 746-1170, 747-1170, 748-1168, 748-1170, 750-883, 752-1170, 753-1170, 755-1170, 756-1162, 756-1170, 760-1170, 763-1049, 763-1170, 764-1170, 765-1007, 765-1010, 766-1170, 767-1170, 768-1170, 769-1170, 770-1170, 771-1170, 772-1170, 774-1012, 774-1020, 774-1055, 775-1170, 776-1170, 778-1170, 779-1142, 780-1061, 782-1170, 784-1168, 784-1170, 785-1053, 785-1170, 787-1170, 788-1170, 789-1170, 790-1170, 791-1067, 791-1075, 792-1170, 793-1170, 794-1170, 795-1025, 795-1036, 795-1170, 796-1170, 797-1052, 797-1170, 798-1170, 804-1170, 807-1114, 808-1170, 810-1170, 811-1170, 812-1170, 813-1031, 814-1170, 817-1170, 818-1170, 819-1154, 823-1038, 823-1074, 823-1089, 824-1170, 829-1075, 830-1170, 838-1113, 838-1124, 838-1170, 839-1170, 840-1170, 841-1170, 842-1109, 843-1098, 843-1170, 846-1170, 847-1096, 847-1170, 849-1170, 850-1170, 852-1015, 853-1170, 854-1167, 855-1170, 857-1170, 858-1170, 859-1170,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	860-1170, 861-1170, 864-1067, 864-1112, 865-1170, 866-1170, 867-1170, 868-1170, 870-1139, 872-1170, 873-1170, 875-1170, 881-1072, 881-1077, 881-1169, 881-1170, 884-1155, 884-1170, 885-1170, 887-1027, 887-1170, 888-1170, 889-974, 893-1170, 895-1156, 895-1170, 897-1170, 898-1170, 900-1170, 901-1068, 904-1170, 905-1170, 906-1170, 907-1027, 907-1170, 912-1157, 912-1170, 915-1170, 918-1164, 921-1170, 922-1167, 922-1170, 924-1161, 929-1047, 929-1170, 930-1170, 932-1170, 933-1122, 933-1170, 935-1170, 936-1170, 942-1170, 946-1170, 949-1149, 950-1170, 965-1170, 967-1170, 983-1170, 987-1170, 992-1127, 998-1139, 998-1170, 1000-1170, 1003-1089, 1006-1170, 1010-1128, 1010-1170, 1019-1170, 1022-1170, 1025-1170, 1026-1170, 1027-1170, 1035-1170, 1046-1170, 1051-1170, 1060-1170, 1071-1170, 1079-1170, 1081-1170, 1105-1170
46/7511057CB1/ 976	1-194, 1-358, 1-414, 1-481, 1-493, 164-753, 343-834, 437-917, 597-974, 637-976
47/7511078CB1/ 1217	1-240, 1-331, 1-542, 1-556, 1-635, 1-648, 167-832, 305-805, 385-1051, 427-986, 469-1157, 514-1203, 666-1217, 680-1217, 986-1217, 1055-1217
48/7511505CB1/ 2520	1-235, 1-283, 1-2520, 4-227, 4-302, 9-266, 24-140, 25-282, 26-264, 26-319, 27-273, 30-315, 38-266, 38-316, 48-297, 64-320, 66-316, 67-231, 67-249, 92-325, 160-702, 325-533, 325-578, 325-919, 325-993, 326-683, 328-638, 346-957, 368-548, 375-552, 379-898, 390-671, 402-562, 403-1002, 434-610, 436-640, 443-1118, 457-1055, 471-910, 472-1059, 474-994, 478-1246, 479-1018, 511-837, 530-742, 535-812, 551-816, 565-1217, 590-1092, 594-1204, 600-735, 614-752, 614-822, 614-1104, 614-1129, 619-1097, 623-864, 623-887, 624-1295, 627-985, 682-1316, 706-1398, 712-1102, 722-1030, 725-1270, 726-1236, 727-958, 728-1240, 728-1245, 731-961, 736-1305, 742-1200, 746-1277, 750-1143, 772-1527, 774-1229, 774-1310, 775-1360, 778-1020, 782-1010, 785-1041, 798-1027, 804-1472, 808-1034, 808-1194, 810-1421, 820-1078, 820-1120, 829-1080, 831-1574, 837-1233, 860-1123, 882-1122, 898-1060, 898-1094, 912-1168, 916-1173, 918-1171, 953-1295, 954-1175, 954-1278, 954-1506, 980-1244, 990-1240, 990-1548, 1005-1283, 1020-1579, 1034-1606, 1056-1354,
	1060-1719, 1061-1315, 1061-1628, 1064-1504, 1072-1400, 1073-1322, 1084-1599, 1090-1664, 1107-1327, 1124-1430, 1127-1934, 1150-1716, 1185-1719, 1194-1442, 1202-1665, 1223-1930, 1229-1485, 1234-1717, 1244-1541, 1262-1379, 1282-1552, 1285-1719, 1291-1533, 1291-1719, 1296-1567, 1296-1588, 1299-1433, 1301-1637, 1319-1581, 1331-1719, 1333-1526, 1335-1563, 1335-1591, 1335-1933, 1378-1719, 1378-2055, 1379-1495, 1394-1605, 1400-1719, 1417-1986, 1421-1708, 1426-1700, 1429-1977, 1431-1869, 1441-1681, 1443-1661, 1448-1702, 1452-1854, 1460-1719, 1463-1719, 1466-1719, 1468-1680, 1468-1686, 1468-1732, 1468-1744, 1468-1966, 1481-2089, 1490-1753, 1490-1819, 1498-1729, 1506-1975, 1507-2073, 1508-1719, 1511-1796, 1513-1779, 1515-2179, 1517-1719, 1517-1747, 1530-1986, 1531-1719, 1547-1864, 1547-2165, 1567-1719, 1574-1763, 1586-1975, 1627-1719, 1629-1885, 1652-1975, 1654-1718, 1654-1719, 1656-1935, 1656-1937, 1656-1938, 1656-1975, 1663-2315, 1680-1963, 1680-1975, 1682-1856, 1682-1933, 1683-1926, 1684-2101, 1687-1949, 1687-1971,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1688-1719, 1689-1985, 1699-2166, 1712-1981, 1712-2255, 1729-1975, 1730-1811, 1733-1975, 1733-2517, 1741-1991, 1743-2172, 1751-1940, 1758-2326, 1759-2017, 1770-2409, 1776-1975, 1784-2277, 1788-2457, 1790-2038, 1790-2044, 1797-2232, 1797-2281, 1820-2447, 1822-2079, 1823-2069, 1823-2256, 1826-2056, 1827-2087, 1851-2112, 1851-2135, 1856-2515, 1858-2477, 1862-2476, 1864-2290, 1865-2187, 1870-2143, 1870-2393, 1890-2465, 1917-2146, 1917-2519, 1922-2479, 1931-2482, 1932-2471, 1948-2194, 1948-2300, 1949-2211, 1949-2214, 1972-2235, 1972-2236, 1973-2487, 1978-2517, 1979-2296, 1979-2520, 1981-2234, 1982-2520, 1988-2343, 1989-2371, 1994-2520, 1998-2520, 2000-2520, 2006-2288, 2010-2274, 2010-2374, 2011-2457, 2012-2482, 2012-2506, 2015-2320, 2017-2520, 2020-2474, 2024-2507, 2029-2325, 2034-2520, 2039-2520, 2044-2520, 2045-2520, 2050-2514, 2051-2479, 2055-2219, 2058-2520, 2062-2520, 2064-2520, 2066-2518, 2067-2505, 2068-2506, 2070-2520, 2072-2520, 2078-2520, 2081-2520, 2082-2520, 2083-2520, 2084-2520, 2085-2517, 2086-2520,
	2095-2518, 2096-2517, 2100-2520, 2102-2342, 2102-2371, 2105-2520, 2109-2520, 2112-2517, 2126-2520, 2127-2520, 2128-2519, 2131-2520, 2135-2520, 2139-2367, 2139-2416, 2151-2520, 2163-2520, 2198-2505, 2198-2518, 2214-2518, 2217-2517, 2222-2520, 2229-2519, 2229-2520, 2234-2520, 2237-2517, 2243-2502, 2272-2520, 2273-2517, 2277-2520, 2285-2515, 2297-2520, 2318-2520, 2319-2520, 2323-2520, 2324-2520, 2325-2510, 2325-2511, 2348-2520, 2357-2480, 2382-2517, 2398-2517, 2404-2520, 2418-2520, 2428-2520, 2459-2520
49/7511552CB1/ 1783	1-63, 1-1749, 6-63, 8-63, 16-63, 181-518, 183-294, 183-312, 183-334, 183-340, 183-358, 183-365, 183-366, 183-367, 183-368, 183-371, 183-375, 183-378, 183-380, 183-384, 183-385, 183-386, 183-390, 183-394, 183-395, 183-399, 183-400, 183-403, 183-404, 183-407, 183-410, 183-412, 183-413, 183-415, 183-416, 183-418, 183-420, 183-421, 183-422, 183-423, 183-427, 183-428, 183-430, 183-432, 183-434, 183-438, 183-439, 183-440, 183-441, 183-445, 183-446, 183-447, 183-448, 183-452, 183-455, 183-456, 183-461, 183-464, 183-465, 183-466, 183-467, 183-468, 183-469, 183-472, 183-473, 183-474, 183-479, 183-480, 183-486, 183-492, 183-493, 183-496, 183-498, 183-519, 183-521, 183-523, 183-548, 183-554, 183-561, 183-562, 183-591, 183-601, 183-602, 183-603, 183-633, 183-636, 183-654, 183-663, 183-673, 183-706, 183-740, 183-754, 183-758, 183-761, 183-762, 183-777, 183-780, 183-810, 183-821, 183-832, 184-566, 184-573, 184-639, 184-663, 184-719, 184-720, 184-728, 184-760, 184-834, 185-800, 186-474, 187-681, 188-401, 188-435, 188-438,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	188-462, 188-744, 189-367, 189-418, 189-671, 189-722, 190-706, 195-443, 195-446, 195-458, 196-408, 196-457, 197-435, 200-764, 204-431, 209-490, 236-827, 240-485, 240-519, 240-623, 258-823, 261-497, 261-552, 269-507, 281-553, 282-523, 286-485, 296-555, 297-536, 297-579, 304-523, 320-611, 321-707, 341-594, 362-979, 363-810, 372-601, 385-658, 397-646, 447-772, 447-774, 447-834, 483-724, 508-833, 525-788, 531-832, 534-722, 536-810, 539-764, 539-783, 539-819, 569-828, 572-831, 607-834, 617-827, 617-830, 670-834, 830-994, 830-1011, 830-1040, 830-1072, 830-1085, 830-1093, 830-1097, 830-1354, 830-1360, 830-1403, 830-1497, 833-1081, 833-1120, 833-1282, 833-1425, 835-1107, 835-1147, 838-1410, 838-1459, 840-1050, 846-1105, 851-1456, 855-1460, 855-1536, 856-1460, 861-1524, 862-1524, 869-1116, 872-1359, 876-1171, 876-1179, 879-1134, 882-1460, 883-1152, 883-1330, 885-1719, 886-1519, 888-1685, 890-1182, 897-1055, 897-1522, 899-1300, 900-1087, 900-1100, 900-1119, 900-1131, 900-1172, 900-1184, 900-1460, 902-1698, 907-1161,
	909-1424, 912-1147, 912-1541, 917-1150, 919-1202, 927-1092, 927-1420, 928-1195, 933-1180, 933-1460, 934-1228, 936-1460, 945-1192, 948-1587, 952-1754, 953-1251, 961-1173, 961-1481, 962-1145, 965-1253, 965-1458, 965-1505, 967-1222, 967-1236, 967-1237, 967-1420, 968-1459, 969-1203, 976-1240, 986-1449, 994-1210, 994-1263, 995-1229, 995-1248, 995-1259, 995-1633, 998-1242, 999-1225, 1001-1252, 1002-1166, 1002-1268, 1003-1271, 1004-1460, 1005-1302, 1006-1606, 1009-1282, 1009-1286, 1009-1291, 1010-1525, 1012-1210, 1013-1456, 1016-1456, 1021-1517, 1023-1457, 1028-1272, 1028-1631, 1030-1408, 1031-1446, 1037-1175, 1037-1244, 1037-1306, 1045-1375, 1047-1320, 1056-1464, 1056-1528, 1063-1294, 1071-1342, 1074-1322, 1074-1495, 1075-1571, 1078-1352, 1079-1463, 1090-1345, 1092-1309, 1104-1354, 1107-1461, 1108-1379, 1114-1406, 1117-1544, 1122-1372, 1131-1363, 1132-1394, 1135-1783, 1138-1395, 1142-1389, 1147-1438, 1147-1451, 1149-1400, 1149-1517, 1149-1735, 1151-1745, 1151-1769, 1159-1715, 1160-1410, 1160-1642, 1167-1551,
	1175-1378, 1175-1432, 1175-1492, 1180-1392, 1187-1466, 1187-1470, 1188-1721, 1189-1446, 1195-1735, 1196-1460, 1196-1742, 1204-1775, 1205-1512, 1208-1476, 1208-1712, 1209-1472, 1214-1715, 1222-1432, 1222-1515, 1222-1578, 1224-1498, 1232-1669, 1234-1519, 1235-1435, 1235-1507, 1241-1537, 1265-1491, 1268-1530, 1269-1459, 1269-1513, 1277-1711, 1278-1543, 1279-1532, 1282-1538, 1289-1733, 1289-1776, 1300-1530, 1300-1538, 1303-1542, 1304-1683, 1307-1707, 1315-1601, 1318-1710, 1321-1746, 1335-1721, 1336-1708, 1338-1438, 1338-1577, 1345-1595, 1353-1761, 1357-1747, 1367-1648, 1379-1588, 1379-1602, 1384-1652, 1384-1664, 1384-1704, 1390-1693, 1392-1733, 1394-1635, 1394-1638, 1400-1692, 1401-1589, 1401-1674, 1404-1598, 1404-1650, 1405-1619, 1405-1628, 1405-1723, 1406-1731, 1410-1649, 1423-1677, 1440-1685, 1443-1596, 1443-1694, 1443-1706, 1457-1735, 1465-1704, 1471-1637, 1473-1732, 1475-1714, 1483-1734, 1488-1751, 1491-1586, 1495-1748, 1528-1688, 1536-1773
50/7511722CB1/ 870	1-243, 1-869, 106-461, 114-414, 115-406, 150-354, 150-358, 150-523, 242-488, 397-870, 500-870, 542-870, 543-825, 662-870, 679-870, 718-869, 736-866, 736-868, 763-869, 800-869

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
51/7511489CB1/ 1572	1-233, 3-272, 8-249, 9-228, 9-271, 18-264, 18-316, 18-1567, 29-626, 32-278, 33-286, 39-598, 42-393, 45-285, 46-339, 51-197, 51-404, 52-506, 53-315, 53-321, 53-337, 53-338, 53-352, 54-259, 55-510, 56-274, 56-838, 58-347, 58-356, 58-364, 59-241, 59-318, 59-324, 59-341, 59-346, 59-363, 59-364, 59-482, 61-243, 61-279, 61-305, 61-477, 61-490, 62-296, 63-360, 64-205, 64-227, 64-299, 64-310, 65-566, 67-596, 68-308, 68-468, 69-248, 69-260, 69-309, 69-313, 69-332, 69-491, 69-499, 70-311, 70-363, 71-382, 71-418, 71-497, 71-508, 72-295, 73-287, 74-359, 76-341, 76-366, 77-331, 77-341, 78-286, 78-351, 78-441, 78-458, 80-388, 80-400, 80-667, 81-217, 81-543, 81-580, 81-617, 81-694, 81-721, 81-722, 82-296, 82-315, 82-339, 82-341, 82-354, 82-376, 82-381, 82-399, 82-402, 82-467, 82-528, 82-550, 82-557, 82-642, 82-728, 83-329, 83-342, 83-350, 83-601, 83-841, 85-294, 85-371, 85-384, 85-402, 86-661, 87-322, 87-337, 87-348, 87-356, 87-427, 87-675, 87-761, 87-765, 88-326, 88-393, 88-823, 89-283, 89-328, 89-350, 89-369, 89-375,
	89-379, 89-381, 89-525, 90-447, 91-356, 91-499, 92-162, 92-525, 93-342, 93-364, 93-460, 93-472, 93-503, 94-348, 94-349, 94-689, 95-341, 96-325, 98-328, 98-375, 100-369, 100-722, 102-343, 102-365, 102-559, 102-570, 102-572, 104-589, 106-355, 106-356, 106-363, 106-364, 106-369, 106-375, 106-376, 106-385, 106-393, 106-395, 106-673, 107-394, 107-409, 108-352, 108-359, 108-361, 108-372, 108-382, 108-439, 108-543, 109-379, 110-601, 114-339, 114-352, 114-511, 115-368, 116-350, 116-419, 117-618, 122-376, 124-627, 125-262, 125-408, 132-380, 132-383, 143-283, 147-399, 147-417, 147-499, 150-533, 162-788, 166-400, 188-461, 198-691, 199-733, 201-479, 201-786, 201-803, 208-430, 210-437, 214-841, 216-689, 216-690, 221-480, 221-492, 234-508, 234-579, 236-498, 236-841, 250-841, 262-720, 263-550, 279-809, 284-745, 289-538, 289-749, 304-801, 305-503, 305-835, 306-622, 316-512, 317-577, 317-604, 327-578, 327-607, 331-575, 332-595, 332-640, 338-638, 339-560, 350-633, 359-625, 361-421, 365-609, 365-762, 372-612, 381-640, 381-660,
	383-597, 383-686, 389-784, 394-682, 397-680, 399-609, 405-597, 407-704, 412-693, 415-703, 419-511, 419-514, 421-592, 430-644, 430-746, 431-695, 432-943, 450-542, 450-714, 455-841, 456-734, 461-709, 461-735, 463-717, 467-736, 467-758, 468-675, 475-729, 482-731, 486-733, 486-773, 490-741, 493-720, 503-749, 505-801, 506-841, 520-827, 525-638, 525-784, 525-841, 526-669, 534-784, 534-798, 534-820, 535-761, 535-789, 546-799, 547-808, 549-751, 553-791, 553-832, 555-810, 556-825, 560-841, 561-837, 569-792, 574-841, 577-826, 585-841, 589-841, 592-841, 593-841, 621-805, 623-820, 711-818, 784-1160, 801-1266, 838-1003, 838-1021, 838-1057, 838-1061, 838-1064, 838-1101, 838-1266, 838-1272, 838-1572, 839-1272, 845-1120, 845-1272, 846-1146, 848-1275, 849-1143, 851-1266, 852-1062, 852-1259, 852-1272, 854-1100, 854-1108, 855-1265, 860-1272, 861-1266, 865-1272, 868-1058, 870-1478, 875-1055, 877-1101, 888-1277, 890-1162, 891-1189, 895-1048, 895-1172, 895-1266, 895-1532, 901-1272, 902-1175, 902-1200, 903-1170, 905-1272,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	906-1437, 906-1567, 914-1360, 914-1402, 918-1174, 924-1174, 928-1160, 931-1211, 932-1232, 932-1492, 932-1528, 935-1466, 936-1320, 938-1169, 939-1161, 942-1185, 946-1272, 947-1272, 949-1218, 955-1204, 955-1557, 957-1182, 957-1504, 966-1199, 966-1210, 972-1252, 972-1567, 975-1251, 975-1265, 977-1530, 980-1272, 984-1326, 988-1309, 990-1266, 992-1305, 993-1189, 1000-1233, 1000-1262, 1002-1275, 1003-1241, 1004-1238, 1004-1344, 1005-1247, 1007-1231, 1008-1214, 1009-1235, 1009-1285, 1010-1184, 1010-1191, 1010-1210, 1010-1230, 1010-1234, 1010-1249, 1010-1251, 1010-1274, 1010-1286, 1010-1287, 1010-1306, 1010-1330, 1013-1232, 1014-1248, 1015-1312, 1017-1486, 1019-1272, 1020-1257, 1026-1572, 1030-1224, 1030-1253, 1030-1313, 1032-1266, 1032-1296, 1032-1572, 1033-1309, 1036-1275, 1041-1553, 1042-1316, 1046-1272, 1049-1362, 1052-1266, 1057-1297, 1058-1272, 1059-1297, 1059-1317, 1059-1365, 1065-1222, 1065-1253, 1065-1256, 1065-1257, 1065-1272, 1065-1278, 1065-1342, 1068-1272, 1069-1509, 1073-1344, 1073-1567, 1074-1351,
	1074-1492, 1077-1272, 1078-1565, 1081-1564, 1083-1567, 1086-1270, 1089-1362, 1091-1570, 1094-1272, 1095-1354, 1095-1570, 1097-1567, 1097-1572, 1098-1572, 1099-1396, 1102-1572, 1103-1344, 1103-1345, 1105-1567, 1105-1572, 1108-1407, 1108-1568, 1110-1567, 1111-1411, 1111-1565, 1111-1572, 1113-1572, 1116-1567, 1117-1569, 1118-1567, 1121-1343, 1121-1371, 1121-1571, 1123-1565, 1123-1567, 1125-1567, 1126-1556, 1127-1572, 1128-1571, 1129-1567, 1130-1567, 1132-1567, 1133-1373, 1141-1572, 1142-1567, 1144-1567, 1144-1572, 1146-1567, 1146-1572, 1147-1391, 1149-1567, 1150-1567, 1151-1272, 1151-1430, 1152-1565, 1154-1572, 1155-1567, 1156-1567, 1157-1567, 1159-1567, 1161-1572, 1162-1572, 1163-1567, 1165-1565, 1165-1569, 1166-1567, 1167-1567, 1171-1567, 1173-1567, 1174-1571, 1179-1423, 1180-1572, 1182-1565, 1184-1272, 1184-1567, 1185-1323, 1185-1567, 1186-1567, 1188-1567, 1196-1452, 1196-1479, 1203-1564, 1204-1457, 1204-1567, 1210-1529, 1211-1381, 1211-1467, 1212-1497, 1213-1565, 1215-1567, 1219-1527, 1219-1567, 1220-1565,
	1221-1567, 1222-1567, 1224-1568, 1225-1566, 1227-1567, 1228-1448, 1233-1477, 1238-1480, 1243-1567, 1246-1567, 1249-1566, 1250-1528, 1250-1558, 1251-1537, 1252-1565, 1256-1567, 1259-1447, 1260-1536, 1260-1559, 1260-1567, 1261-1565, 1262-1564, 1267-1567, 1268-1507, 1272-1567, 1277-1532, 1285-1567, 1303-1565, 1305-1567, 1310-1447, 1310-1453, 1318-1572, 1319-1567, 1321-1558, 1322-1566, 1324-1506, 1324-1542, 1324-1565, 1325-1567, 1326-1565, 1337-1461, 1338-1567, 1340-1529, 1340-1551, 1341-1561, 1359-1568, 1362-1567, 1369-1567, 1398-1516, 1400-1572, 1405-1567, 1412-1502, 1429-1567, 1438-1567, 1506-1567

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
52/7511497CB1/ 1042	1-236, 1-241, 1-252, 1-288, 2-218, 2-233, 2-237, 2-244, 2-266, 2-269, 2-275, 2-291, 2-304, 8-294, 11-186, 12-1039, 22-247, 22-249, 22-287, 22-291, 22-305, 23-305, 24-277, 24-278, 24-281, 24-282, 24-284, 24-290, 24-296, 24-305, 25-266, 25-274, 25-281, 25-285, 25-301, 25-302, 26-225, 26-253, 26-279, 26-294, 26-302, 26-303, 26-305, 27-265, 27-275, 27-297, 27-301, 27-305, 28-287, 28-295, 28-303, 28-305, 29-273, 29-280, 29-290, 29-305, 30-304, 34-263, 34-292, 35-248, 35-274, 35-288, 35-294, 35-296, 35-305, 36-138, 36-255, 36-271, 36-305, 37-276, 37-282, 37-296, 37-297, 38-247, 39-250, 39-282, 39-287, 39-305, 41-258, 41-299, 41-305, 42-276, 42-305, 43-190, 43-221, 43-267, 43-284, 43-305, 44-270, 44-274, 44-284, 45-231, 45-282, 46-285, 47-305, 49-275, 49-296, 49-305, 49-340, 52-301, 57-272, 71-197, 77-163, 129-305, 307-546, 322-864, 324-971, 329-617, 333-573, 341-679, 342-638, 342-906, 343-836, 348-602, 350-595, 353-506, 354-600, 354-612, 355-546, 356-721, 364-536, 366-782, 386-696, 390-900, 392-604, 400-961,
	408-670, 409-660, 410-686, 417-990, 421-638, 429-640, 440-615, 461-1042, 465-659, 465-666, 465-718, 465-751, 470-1039, 479-747, 479-817, 479-938, 483-698, 484-684, 490-745, 505-908, 505-1042, 506-725, 506-788, 511-1042, 519-898, 527-1042, 532-1042, 534-804, 535-789, 562-816, 562-833, 562-840, 565-1034, 576-900, 580-900, 587-837, 587-900, 588-864, 591-900, 593-843, 593-866, 593-867, 593-878, 594-884, 597-909, 611-858, 611-900, 616-900, 618-889, 618-900, 621-900, 625-1042, 628-900, 631-900, 639-806, 643-900, 644-900, 645-900, 647-1042, 648-900, 648-1042, 649-900, 659-1042, 663-900, 665-916, 665-918, 666-900, 680-963, 681-954, 682-1042, 686-1042, 691-900, 708-900, 708-918, 711-881, 711-900, 722-900, 722-986, 723-1012, 742-900, 744-899, 745-917, 761-976, 761-993, 763-900, 772-900, 780-900, 780-1001, 780-1042, 794-1006, 810-900, 813-920, 813-1042, 814-1042, 815-900, 815-1042, 824-900, 878-1042
53/7511498CB1/ 921	1-236, 1-241, 1-252, 1-288, 1-347, 1-385, 1-486, 1-683, 2-218, 2-233, 2-237, 2-244, 2-266, 2-269, 2-275, 2-291, 2-320, 3-313, 3-340, 4-471, 8-294, 11-186, 12-921, 19-638, 22-247, 22-249, 22-287, 22-291, 22-311, 22-570, 22-582, 22-683, 23-321, 24-277, 24-278, 24-281, 24-282, 24-284, 24-290, 24-296, 24-327, 24-496, 24-530, 25-266, 25-274, 25-281, 25-285, 25-301, 25-302, 25-311, 25-327, 25-376, 25-508, 25-517, 25-549, 25-558, 26-225, 26-253, 26-279, 26-294, 26-302, 26-303, 26-308, 26-310, 26-312, 26-595, 26-673, 27-265, 27-275, 27-297, 27-301, 27-316, 28-287, 28-303, 28-310, 28-314, 28-316, 28-320, 28-322, 28-325, 28-334, 28-339, 28-392, 28-614, 29-273, 29-280, 29-290, 29-316, 29-637, 30-304, 32-340, 33-429, 34-263, 34-292, 35-248, 35-274, 35-288, 35-294, 35-296, 35-313, 35-320, 35-592, 35-602, 36-138, 36-255, 36-271, 36-325, 36-439, 37-276, 37-282, 37-296, 37-297, 37-337, 37-553, 37-614, 37-627, 37-683, 38-247, 39-250, 39-282, 39-287, 39-313, 39-314, 39-323, 41-258, 41-299, 41-305, 41-310, 41-318, 41-337,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	41-554, 41-575, 42-276, 42-310, 42-615, 43-190, 43-221, 43-267, 43-284, 43-305, 43-322, 43-324, 43-330, 43-331, 43-387, 43-513, 43-515, 43-605, 44-270, 44-274, 44-284, 45-231, 45-282, 45-337, 46-285, 47-480, 47-589, 49-275, 49-296, 49-315, 50-333, 51-499, 52-301, 52-436, 53-657, 54-395, 57-272, 57-675, 58-445, 71-197, 71-356, 77-163, 109-385, 109-683, 116-415, 119-349, 129-381, 138-400, 143-414, 147-538, 153-517, 160-432, 174-538, 190-432, 191-441, 192-401, 196-407, 198-389, 199-419, 231-621, 246-448, 253-474, 265-604, 284-385, 300-583, 307-517, 313-471, 315-604, 327-623, 345-683, 358-604, 358-617, 367-610, 376-524, 386-667, 404-643, 430-670, 445-683, 447-683, 450-603, 451-683, 452-643, 461-633, 468-921, 483-921, 489-683, 576-921, 683-921, 687-921, 718-921, 719-921, 760-921, 801-921
54/7511612CB1/ 1754	1-294, 6-374, 9-668, 11-229, 11-303, 11-467, 11-502, 11-1677, 16-226, 23-305, 39-132, 62-222, 62-309, 62-345, 62-496, 65-250, 65-337, 66-326, 67-304, 73-319, 73-506, 74-624, 77-336, 78-329, 80-328, 84-371, 85-403, 86-286, 135-389, 147-330, 168-424, 185-416, 232-489, 256-501, 257-498, 341-877, 363-668, 413-877, 414-700, 461-698, 473-605, 505-760, 505-1128, 507-648, 507-789, 508-706, 508-715, 509-816, 513-1084, 518-772, 534-785, 541-1026, 543-977, 543-1057, 548-1077, 550-1054, 550-1076, 553-1040, 557-830, 565-1174, 566-814, 566-992, 567-812, 567-833, 567-836, 572-845, 579-863, 583-847, 588-1249, 589-777, 589-1180, 598-1145, 601-873, 612-847, 618-1053, 621-900, 623-897, 626-1101, 631-902, 636-880, 636-890, 636-893, 636-1242, 646-872, 670-928, 673-1264, 673-1390, 685-959, 687-956, 687-970, 696-971, 703-972, 707-1360, 738-1141, 740-871, 740-1012, 742-1034, 744-987, 769-1020, 779-1043, 780-1427, 781-1063, 781-1431, 784-1224, 791-975, 793-1068, 800-1217, 801-857, 801-1100, 811-1059, 831-1031, 831-1085, 831-1118,
	860-1414, 861-1084, 862-1538, 871-1113, 874-1679, 880-1171, 881-1145, 886-1130, 898-1464, 900-1091, 900-1476, 902-1543, 902-1663, 903-1142, 907-1217, 907-1232, 917-1161, 917-1167, 918-1188, 921-1217, 923-1595, 929-1180, 938-1193, 944-1200, 944-1208, 944-1232, 944-1431, 946-1163, 960-1206, 965-1664, 967-1683, 974-1519, 982-1403, 993-1544, 993-1552, 1009-1510, 1011-1286, 1036-1641, 1038-1285, 1040-1300, 1049-1638, 1059-1495, 1060-1484, 1063-1530, 1066-1331, 1075-1281, 1075-1364, 1075-1570, 1082-1299, 1084-1352, 1084-1379, 1093-1519, 1094-1489, 1097-1351, 1097-1368, 1098-1213, 1099-1549, 1107-1583, 1114-1563, 1114-1566, 1114-1568, 1114-1576, 1115-1684, 1123-1624, 1128-1404, 1135-1379, 1135-1616, 1137-1675, 1143-1680, 1152-1685, 1153-1474, 1155-1423, 1160-1407, 1161-1645, 1166-1413, 1167-1682, 1168-1614, 1176-1424, 1180-1415, 1180-1568, 1181-1682, 1184-1682, 1188-1628, 1192-1666, 1194-1360, 1194-1444, 1194-1488, 1196-1668, 1199-1680, 1204-1666, 1205-1665, 1206-1571, 1206-1669, 1208-1675, 1209-1666, 1210-1685,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1215-1665, 1218-1665, 1221-1670, 1227-1670, 1231-1670, 1234-1665, 1235-1696, 1236-1665, 1237-1666, 1237-1668, 1240-1666, 1245-1424, 1246-1666, 1246-1668, 1248-1676, 1249-1667, 1251-1456, 1251-1548, 1252-1665, 1253-1430, 1256-1665, 1256-1666, 1258-1666, 1259-1431, 1261-1582, 1261-1668, 1262-1664, 1263-1671, 1264-1498, 1264-1667, 1268-1569, 1268-1666, 1269-1663, 1272-1526, 1274-1669, 1282-1666, 1284-1665, 1285-1614, 1288-1672, 1289-1667, 1293-1567, 1298-1680, 1299-1666, 1308-1665, 1309-1628, 1310-1666, 1312-1665, 1313-1614, 1314-1590, 1321-1683, 1323-1669, 1323-1674, 1335-1561, 1335-1666, 1336-1668, 1337-1734, 1348-1568, 1394-1698, 1402-1568, 1406-1666, 1407-1667, 1411-1671, 1413-1666, 1414-1665, 1416-1665, 1417-1668, 1424-1672, 1436-1665, 1440-1623, 1454-1662, 1454-1665, 1457-1696, 1468-1665, 1471-1667, 1500-1754, 1570-1686, 1585-1681
55/7511624CB1/ 2262	1-248, 1-266, 1-269, 1-289, 1-521, 1-2262, 5-306, 6-301, 11-480, 11-879, 12-184, 12-212, 12-384, 13-346, 14-525, 15-288, 15-406, 16-282, 17-425, 20-149, 20-339, 21-247, 21-250, 21-260, 22-278, 23-310, 23-372, 26-456, 29-258, 29-278, 29-309, 31-225, 31-267, 31-271, 31-276, 31-525, 34-315, 34-520, 35-102, 35-259, 47-379, 50-280, 57-501, 58-332, 62-260, 75-368, 91-327, 111-518, 140-395, 150-437, 152-525, 154-438, 165-388, 173-299, 219-476, 219-503, 268-512, 301-503, 522-706, 522-777, 522-1103, 522-1131, 522-1135, 522-1142, 522-1145, 522-1179, 522-1293, 526-834, 526-1175, 538-833, 538-1015, 540-820, 541-807, 541-1154, 542-1286, 581-897, 587-1039, 598-1051, 628-807, 628-1003, 628-1036, 628-1058, 633-1060, 654-899, 656-1143, 677-1246, 684-968, 684-1188, 692-945, 693-890, 701-1280, 705-1296, 715-1405, 723-984, 724-981, 725-865, 726-1390, 727-1008, 735-1176, 744-1028, 745-982, 746-1395, 747-1383, 747-1390, 754-1405, 758-914, 760-916, 761-1038, 776-1291, 777-1424, 780-1427, 782-1002, 786-1068, 789-1050, 789-1276,
	792-1433, 793-1578, 799-1086, 808-1094, 811-1057, 829-1643, 847-1099, 847-1138, 847-1578, 857-1126, 857-1130, 858-1098, 863-1578, 867-1022, 867-1138, 870-1244, 871-1136, 871-1141, 872-1165, 880-1137, 882-1578, 887-1700, 897-1135, 898-1091, 904-1433, 916-1220, 917-1130, 919-1149, 935-1070, 936-1050, 940-1164, 943-1186, 949-1227, 949-1336, 950-1187, 960-1218, 965-1126, 965-1262, 965-1328, 965-1376, 965-1401, 965-1518, 965-1718, 971-1270, 978-1238, 996-1236, 996-1281, 996-1397, 996-1573, 1004-1274, 1007-1572, 1012-1612, 1013-1578, 1015-1262, 1017-1331, 1017-1603, 1017-1712, 1020-1280, 1025-1263, 1028-1271, 1034-1222, 1035-1491, 1042-1639, 1044-1339, 1049-1578, 1051-1548, 1057-1303, 1058-1372, 1058-1476, 1059-1345, 1066-1707, 1070-1401, 1071-1973, 1075-1313, 1078-1534, 1085-1324, 1085-1354, 1089-1683, 1092-1222, 1094-1364, 1095-1371, 1096-1428, 1100-1427, 1102-1304, 1102-1391, 1103-1713, 1105-1328, 1110-1715, 1113-1906, 1115-1443, 1117-1317, 1117-1435, 1126-1539, 1126-1973, 1130-1402, 1142-1415, 1145-1637,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1158-1396, 1159-1973, 1189-1973, 1190-1513, 1194-1973, 1196-1973, 1199-1969, 1205-1479, 1205-1853, 1220-1496, 1220-1499, 1224-1878, 1226-1471, 1226-1511, 1231-1712, 1232-1492, 1237-1426, 1237-1487, 1237-1516, 1243-1516, 1247-1494, 1253-1796, 1258-1953, 1264-1518, 1264-1524, 1266-1852, 1272-1476, 1272-1524, 1273-1531, 1278-1668, 1279-1973, 1289-1770, 1290-1585, 1291-1564, 1294-1552, 1300-1970, 1302-1557, 1305-1572, 1306-1763, 1307-1894, 1322-1600, 1327-1575, 1327-1861, 1329-1429, 1332-1701, 1334-1585, 1336-1760, 1340-1986, 1341-1791, 1345-1983, 1346-1805, 1356-1624, 1359-1733, 1359-1734, 1361-1643, 1361-1819, 1362-1973, 1365-1616, 1365-1625, 1367-1566, 1367-1612, 1368-1594, 1371-1620, 1372-1608, 1373-1510, 1373-1969, 1378-1595, 1380-1612, 1381-1664, 1384-1642, 1385-1686, 1389-1800, 1389-1845, 1392-1700, 1392-1864, 1393-1574, 1393-2074, 1396-1884, 1397-1658, 1397-1679, 1398-1715, 1399-1715, 1399-1732, 1400-2107, 1403-1562, 1406-2040, 1406-2089, 1407-1695, 1407-1708, 1416-1931, 1417-1917, 1418-1704, 1419-1748,
	1428-1913, 1434-1668, 1439-1771, 1448-1702, 1449-1875, 1473-2167, 1479-1885, 1480-1753, 1480-1792, 1482-1728, 1482-1744, 1483-1801, 1490-1767, 1492-2068, 1492-2074, 1493-1759, 1493-1799, 1493-1985, 1494-2180, 1502-1711, 1504-2221, 1507-1767, 1510-2184, 1511-1766, 1519-1677, 1527-1792, 1527-2112, 1531-1963, 1536-1829, 1536-1830, 1536-1943, 1539-1765, 1546-1785, 1546-1788, 1546-1811, 1548-1702, 1548-1856, 1555-1775, 1555-1818, 1555-2041, 1557-1797, 1557-2069, 1557-2115, 1558-1821, 1558-2013, 1558-2249, 1559-1820, 1560-1875, 1560-1985, 1562-2098, 1566-1867, 1567-1617, 1567-1673, 1567-1812, 1567-1861, 1567-1867, 1567-1906, 1567-2063, 1567-2098, 1567-2187, 1570-1921, 1601-2208, 1602-1860, 1602-1869, 1603-2208, 1606-1919, 1613-2192, 1616-1836, 1624-1854, 1624-2003, 1626-2262, 1629-2063, 1636-2235, 1646-1880, 1646-1894, 1646-1897, 1646-1901, 1648-2256, 1648-2262, 1650-2031, 1652-2207, 1655-1879, 1655-1907, 1656-1929, 1657-2071, 1657-2254, 1657-2262, 1660-2217, 1663-2185, 1664-2262, 1676-1952, 1684-2243, 1687-1970,
	1693-2262, 1700-2226, 1707-2198, 1707-2201, 1707-2202, 1707-2208, 1707-2254, 1707-2257, 1707-2262, 1708-2139, 1717-1899, 1717-2253, 1717-2262, 1722-1997, 1724-2215, 1725-2262, 1726-2054, 1726-2097, 1726-2142, 1728-1911, 1728-1953, 1732-2088, 1735-2041, 1737-2262, 1740-2262, 1743-2262, 1748-1934, 1748-1984, 1749-2253, 1755-2258, 1759-2262, 1760-2262, 1761-2262, 1764-2215, 1765-1978, 1768-2010, 1772-2029, 1775-2221, 1775-2257, 1776-2257, 1776-2262, 1778-2262, 1783-2262, 1785-2257, 1787-2262, 1788-2229, 1790-2051, 1790-2256, 1793-2262, 1794-2258, 1796-2045, 1801-2257, 1803-2262, 1805-2258, 1805-2262, 1806-2262, 1807-2262, 1808-2229, 1808-2257, 1808-2262, 1810-2258, 1811-2254, 1812-2121, 1812-2259, 1814-2262, 1815-2054, 1815-2262, 1816-2220, 1816-2262, 1817-2262, 1819-2262, 1821-2220, 1821-2254, 1821-2262, 1822-2259, 1827-2257, 1828-2219, 1828-2258, 1831-2259, 1832-2093, 1834-2257, 1835-2262, 1836-2080, 1836-2113, 1836-2219, 1836-2257, 1837-2244, 1839-2262, 1841-2257, 1842-2262, 1844-2231, 1844-2246, 1844-2257,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1845-2257, 1845-2258, 1846-1874, 1846-2257, 1846-2259, 1846-2262, 1847-2257, 1847-2258, 1850-2257, 1851-2218, 1854-2091, 1854-2216, 1854-2257, 1855-2257, 1856-2256, 1856-2257, 1856-2262, 1859-2143, 1859-2253, 1860-2257, 1861-2244, 1861-2262, 1862-2257, 1864-2259, 1867-2233, 1867-2257, 1867-2262, 1868-2231, 1868-2262, 1870-2080, 1870-2257, 1871-2229, 1871-2262, 1872-2262, 1875-2262, 1876-2258, 1878-2162, 1878-2257, 1879-2262, 1882-2159, 1884-2257, 1884-2258, 1885-2256, 1885-2258, 1888-2161, 1890-2260, 1893-2261, 1893-2262, 1895-2163, 1896-2236, 1897-2262, 1906-2237, 1908-2262, 1909-2167, 1909-2259, 1910-2132, 1912-2074, 1912-2100, 1912-2175, 1913-2259, 1914-2168, 1916-2262, 1919-2191, 1925-2177, 1928-2154, 1928-2217, 1928-2257, 1931-2260, 1932-2174, 1932-2231, 1935-2257, 1939-2157, 1939-2258, 1942-2083, 1945-2257, 1954-2261, 1956-2262, 1959-2258, 1965-2254, 1966-2225, 1970-2262, 1974-2257, 1975-2001, 1979-2257, 1983-2257, 1983-2259, 1986-2257, 1993-2257, 1994-2262, 1996-2213, 1999-2256, 2006-2257, 2008-2258,
	2008-2262, 2009-2257, 2012-2257, 2024-2257, 2026-2257, 2026-2262, 2027-2249, 2027-2262, 2029-2257, 2032-2260, 2033-2257, 2045-2257, 2046-2215, 2053-2262, 2060-2256, 2066-2262, 2070-2257, 2075-2257, 2075-2262, 2083-2262, 2085-2216, 2097-2262, 2100-2261, 2102-2259, 2104-2257, 2105-2220, 2105-2257, 2138-2262, 2154-2262, 2155-2262, 2156-2262, 2172-2258, 2182-2262, 2188-2258
56/7511626CB1/ 2295	1-248, 1-266, 1-269, 1-289, 1-2295, 5-306, 6-301, 12-184, 12-212, 13-332, 15-263, 15-288, 16-282, 17-332, 20-149, 20-332, 21-247, 21-250, 21-260, 22-278, 23-310, 23-330, 26-332, 29-258, 29-278, 29-309, 31-225, 31-267, 31-271, 31-276, 34-315, 34-332, 35-102, 35-259, 50-280, 58-355, 62-260, 91-327, 165-332, 173-299, 332-614, 332-1180, 338-604, 366-942, 366-1003, 382-631, 382-638, 382-917, 405-886, 423-1143, 431-946, 434-673, 444-685, 448-662, 450-670, 450-723, 467-726, 467-1173, 495-734, 501-1065, 517-814, 517-885, 528-810, 532-1175, 535-1164, 535-1168, 535-1178, 541-1212, 544-1136, 550-739, 555-1326, 559-867, 559-1208, 571-866, 571-1048, 573-853, 574-840, 574-1187, 575-1319, 614-930, 620-1072, 631-1084, 661-840, 661-1036, 661-1069, 661-1091, 666-1093, 687-932, 689-1176, 710-1279, 717-1001, 717-1221, 725-978, 726-923, 734-1052, 734-1313, 738-1329, 748-1438, 756-1017, 757-1014, 758-898, 759-1423, 760-1041, 768-1209, 779-1428, 780-1416, 780-1423, 787-1438, 791-947, 793-949, 794-1071, 809-1324, 810-1457,
	813-1460, 815-1035, 819-1101, 822-1083, 822-1309, 825-1466, 826-1611, 832-1072, 832-1119, 841-1127, 844-1090, 862-1676, 880-1132, 880-1171, 880-1611, 890-1159, 890-1163, 891-1131, 896-1611, 900-1055, 900-1171, 903-1277, 904-1169, 904-1174, 905-1198, 913-1170, 915-1611, 920-1733, 930-1168, 931-1124, 937-1466, 949-1253, 950-1163, 952-1182, 968-1103, 969-1083, 973-1197, 976-1219, 982-1260, 982-1369, 983-1220, 993-1251, 998-1159, 998-1295, 998-1361, 998-1409, 998-1434, 998-1551, 998-1751, 1004-1303, 1011-1271, 1029-1269, 1029-1314, 1029-1430, 1029-1606, 1037-1307, 1040-1605, 1045-1645, 1046-1611, 1048-1295, 1050-1364, 1050-1636, 1050-1745, 1053-1313, 1058-1296, 1061-1304, 1067-1255, 1068-1524, 1075-1672, 1077-1372, 1082-1611, 1084-1581, 1090-1336, 1091-1405, 1091-1509, 1092-1378, 1099-1740, 1103-1434, 1104-2006, 1108-1346, 1111-1567, 1118-1357, 1118-1387, 1122-1716, 1125-1255, 1127-1397, 1128-1404, 1129-1461, 1133-1460, 1135-1337, 1135-1424, 1136-1746, 1138-1361, 1143-1748, 1146-1939, 1148-1476, 1150-1350,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1150-1468, 1159-1572, 1159-2006, 1163-1435, 1175-1448, 1178-1670, 1191-1429, 1192-2006, 1222-2006, 1223-1546, 1227-2006, 1229-2006, 1232-2002, 1238-1512, 1238-1886, 1253-1529, 1253-1532, 1257-1911, 1259-1504, 1259-1544, 1264-1745, 1265-1525, 1270-1459, 1270-1520, 1270-1549, 1276-1549, 1280-1527, 1286-1829, 1291-1986, 1297-1551, 1297-1557, 1299-1885, 1305-1509, 1305-1557, 1306-1564, 1311-1701, 1312-2006, 1322-1803, 1323-1618, 1324-1597, 1327-1585, 1333-2003, 1335-1590, 1338-1605, 1339-1796, 1340-1927, 1355-1633, 1360-1608, 1360-1894, 1362-1462, 1365-1734, 1367-1618, 1369-1793, 1373-2019, 1374-1824, 1378-2016, 1379-1838, 1389-1657, 1392-1766, 1392-1767, 1394-1676, 1394-1852, 1395-2006, 1398-1649, 1398-1658, 1400-1599, 1400-1645, 1401-1627, 1404-1653, 1405-1641, 1406-1543, 1406-2002, 1411-1628, 1413-1645, 1414-1697, 1417-1675, 1418-1719, 1422-1833, 1422-1878, 1425-1733, 1425-1897, 1426-1607, 1426-2107, 1429-1917, 1430-1691, 1430-1712, 1431-1748, 1432-1748, 1432-1765, 1433-2140, 1436-1595, 1439-2073, 1439-2122,
	1440-1728, 1440-1741, 1449-1964, 1450-1950, 1451-1737, 1452-1781, 1461-1946, 1467-1701, 1472-1804, 1481-1735, 1482-1908, 1506-2200, 1512-1918, 1513-1786, 1513-1825, 1515-1761, 1515-1777, 1516-1834, 1523-1800, 1525-2101, 1525-2107, 1526-1792, 1526-1832, 1526-2018, 1527-2213, 1535-1744, 1537-2254, 1540-1800, 1543-2217, 1544-1799, 1552-1710, 1560-1825, 1560-2145, 1564-1996, 1569-1862, 1569-1863, 1569-1976, 1572-1798, 1579-1818, 1579-1821, 1579-1844, 1581-1735, 1581-1889, 1588-1808, 1588-1851, 1588-2074, 1590-1830, 1590-2102, 1590-2148, 1591-1854, 1591-2046, 1591-2282, 1592-1853, 1593-1908, 1593-2018, 1595-2131, 1599-1900, 1600-1650, 1600-1706, 1600-1845, 1600-1894, 1600-1900, 1600-1939, 1600-2096, 1600-2131, 1600-2220, 1603-1954, 1634-2241, 1635-1893, 1635-1902, 1636-2241, 1639-1952, 1646-2225, 1649-1869, 1657-1887, 1657-2036, 1659-2295, 1662-2096, 1669-2268, 1679-1913, 1679-1927, 1679-1930, 1679-1934, 1681-2289, 1681-2295, 1683-2064, 1685-2240, 1688-1912, 1688-1940, 1689-1962, 1690-2104, 1690-2287, 1690-2295,
	1693-2250, 1696-2218, 1697-2295, 1709-1985, 1717-2276, 1720-2003, 1726-2295, 1733-2259, 1740-2231, 1740-2234, 1740-2235, 1740-2241, 1740-2287, 1740-2290, 1740-2295, 1741-2172, 1750-1932, 1750-2286, 1750-2295, 1755-2030, 1757-2248, 1758-2295, 1759-2087, 1759-2130, 1759-2175, 1761-1944, 1761-1986, 1765-2121, 1768-2074, 1770-2295, 1773-2295, 1776-2295, 1781-1967, 1781-2017, 1782-2286, 1788-2291, 1792-2295, 1793-2295, 1794-2295, 1797-2248, 1798-2011, 1801-2043, 1805-2062, 1808-2254, 1808-2290, 1809-2290, 1809-2295, 1811-2295, 1816-2295, 1818-2290, 1820-2295, 1821-2262, 1823-2084, 1823-2289, 1826-2295, 1827-2291, 1829-2078, 1834-2290, 1836-2295, 1838-2291, 1838-2295, 1839-2295, 1840-2295, 1841-2262, 1841-2290, 1841-2295, 1843-2291, 1844-2287, 1845-2154, 1845-2292, 1847-2295, 1848-2087, 1848-2295, 1849-2253, 1849-2295, 1850-2295, 1852-2295, 1854-2253, 1854-2287, 1854-2295, 1855-2292, 1860-2290, 1861-2252, 1861-2291, 1864-2292, 1865-2126, 1867-2290, 1868-2295, 1869-2113, 1869-2146, 1869-2252, 1869-2290, 1870-2277,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1872-2295, 1874-2290, 1875-2295, 1877-2264, 1877-2279, 1877-2290, 1878-2290, 1878-2291, 1879-1907, 1879-2290, 1879-2292, 1879-2295, 1880-2290, 1880-2291, 1883-2290, 1884-2251, 1887-2124, 1887-2249, 1887-2290, 1888-2290, 1889-2289, 1889-2290, 1889-2295, 1892-2176, 1892-2286, 1893-2290, 1894-2277, 1894-2295, 1895-2290, 1897-2292, 1900-2266, 1900-2290, 1900-2295, 1901-2264, 1901-2295, 1903-2113, 1903-2290, 1904-2262, 1904-2295, 1905-2295, 1908-2295, 1909-2291, 1911-2195, 1911-2290, 1912-2295, 1915-2192, 1917-2290, 1917-2291, 1918-2289, 1918-2291, 1921-2194, 1923-2293, 1926-2294, 1926-2295, 1928-2196, 1929-2269, 1930-2295, 1939-2270, 1941-2295, 1942-2200, 1942-2292, 1943-2165, 1945-2107, 1945-2133, 1945-2208, 1946-2292, 1947-2201, 1949-2295, 1952-2224, 1958-2210, 1961-2187, 1961-2250, 1961-2290, 1964-2293, 1965-2207, 1965-2264, 1968-2290, 1972-2190, 1972-2291, 1975-2116, 1978-2290, 1987-2294, 1989-2295, 1992-2291, 1998-2287, 1999-2258, 2003-2295, 2007-2290, 2008-2034, 2012-2290, 2016-2290, 2016-2292, 2019-2290,
	2026-2290, 2027-2295, 2029-2246, 2032-2289, 2039-2290, 2041-2291, 2041-2295, 2042-2290, 2045-2290, 2057-2290, 2059-2290, 2059-2295, 2060-2282, 2060-2295, 2062-2290, 2065-2293, 2066-2290, 2078-2290, 2079-2248, 2086-2295, 2093-2289, 2099-2295, 2103-2290, 2108-2290, 2108-2295, 2116-2295, 2118-2249, 2130-2295, 2133-2294, 2135-2292, 2137-2290, 2138-2253, 2138-2290, 2171-2295, 2187-2295, 2188-2295, 2189-2295, 2205-2291, 2215-2295, 2221-2291
57/7512885CB1/ 1108	1-32, 1-215, 1-223, 1-245, 1-252, 1-261, 1-265, 1-273, 1-274, 1-283, 1-303, 1-344, 1-363, 1-423, 1-425, 1-432, 1-447, 1-451, 1-481, 1-497, 1-525, 1-535, 1-539, 1-589, 1-591, 1-594, 1-1108, 2-175, 4-32, 5-291, 8-283, 12-836, 13-284, 14-313, 14-627, 15-264, 15-281, 17-256, 17-312, 18-200, 18-218, 18-285, 18-286, 18-305, 20-306, 21-228, 21-235, 21-275, 21-295, 21-455, 22-602, 23-267, 23-340, 27-302, 28-403, 28-489, 28-498, 28-516, 29-275, 29-295, 29-296, 29-308, 29-309, 29-327, 29-440, 30-291, 30-460, 30-600, 31-220, 31-277, 31-285, 31-299, 31-321, 31-343, 31-545, 31-555, 32-416, 32-627, 32-630, 37-289, 37-313, 38-456, 40-368, 42-368, 42-384, 44-290, 56-298, 57-349, 58-501, 127-598, 133-556, 177-623, 182-437, 211-469, 219-478, 234-630, 235-515, 268-488, 337-552, 349-545, 350-627, 353-861, 368-566, 403-566, 423-621, 627-1086, 696-1101, 827-1103, 890-1039, 928-1101, 937-1056, 937-1065, 937-1094, 948-1101, 949-1102, 955-1108
58/7511965CB1/ 1170	1-603, 1-639, 1-676, 45-592, 271-658, 289-386, 289-403, 289-1170, 325-1170, 344-614, 459-717, 474-671, 716-829, 733-991, 769-1015, 782-1047, 818-1083, 834-1088, 856-1107

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
59/7512403CB1/ 2506	1-524, 7-674, 7-678, 13-257, 13-2504, 19-271, 21-243, 23-276, 23-317, 23-339, 34-300, 34-677, 35-349, 37-291, 42-316, 49-303, 64-678, 65-678, 89-372, 89-540, 89-678, 98-620, 107-678, 126-675, 168-439, 181-804, 183-494, 194-462, 215-463, 252-464, 273-458, 273-517, 361-596, 362-678, 377-573, 675-909, 675-956, 675-967, 675-968, 676-1144, 690-965, 708-952, 708-963, 708-972, 715-994, 715-1005, 715-1302, 758-979, 769-1010, 770-1060, 771-1027, 771-1049, 772-1390, 795-1364, 795-1411, 799-1100, 807-1322, 827-1312, 841-1266, 845-1074, 846-1057, 846-1115, 862-1267, 866-1162, 877-1056, 884-1054, 897-1035, 910-1131, 910-1505, 916-1154, 931-1322, 946-1189, 950-1149, 968-1423, 973-1459, 991-1230, 992-1246, 992-1271, 998-1428, 1007-1282, 1013-1565, 1023-1508, 1025-1309, 1032-1296, 1032-1386, 1042-1511, 1045-1506, 1050-1506, 1052-1508, 1056-1423, 1057-1506, 1060-1508, 1062-1511, 1063-1369, 1066-1510, 1079-1510, 1111-1328, 1115-1511, 1121-1508, 1126-1506, 1137-1508, 1140-1424, 1144-1505, 1152-1511, 1155-1508, 1162-1507,
	1173-1624, 1173-1629, 1179-1659, 1180-1437, 1187-1484, 1204-1314, 1204-1446, 1204-1464, 1204-1508, 1204-1801, 1206-1458, 1209-1478, 1209-1506, 1213-1459, 1216-1506, 1218-1666, 1219-1511, 1220-1506, 1222-1507, 1230-2000, 1234-1506, 1239-2087, 1240-2087, 1241-1506, 1255-1511, 1261-1506, 1261-1779, 1262-1836, 1265-1763, 1307-1428, 1312-1606, 1320-1890, 1321-2094, 1376-1904, 1377-2094, 1381-1542, 1398-1646, 1399-1743, 1399-1920, 1403-1944, 1407-1641, 1417-2091, 1429-2092, 1442-2023, 1447-2086, 1468-2110, 1472-1728, 1478-2092, 1518-1739, 1540-1890, 1546-1802, 1577-2177, 1621-1882, 1621-2119, 1622-1878, 1625-2051, 1631-2063, 1632-1890, 1632-2194, 1648-1815, 1676-1977, 1676-2103, 1693-2352, 1700-1949, 1700-2380, 1704-2483, 1705-2216, 1705-2247, 1705-2254, 1705-2268, 1705-2291, 1705-2319, 1705-2324, 1705-2329, 1705-2339, 1705-2340, 1705-2349, 1705-2362, 1705-2364, 1705-2366, 1705-2404, 1705-2421, 1705-2424, 1705-2429, 1705-2447, 1705-2463, 1715-1995, 1718-2489, 1720-1975, 1721-1940, 1729-1946, 1739-2035, 1757-2402,
	1766-2463, 1772-2362, 1780-2363, 1785-2123, 1791-2409, 1805-2088, 1805-2369, 1806-2433, 1809-2098, 1830-2103, 1835-2082, 1835-2103, 1838-2104, 1839-2103, 1839-2123, 1840-2159, 1854-2123, 1855-2378, 1855-2407, 1862-2485, 1876-2438, 1891-2168, 1925-2123, 1935-2447, 1938-2423, 1941-2074, 1954-2506, 1959-2089, 1959-2122, 1966-2094, 1966-2113, 1966-2200, 1966-2253, 1966-2493, 1967-2229, 1976-2264, 1977-2216, 1977-2239, 2005-2103, 2006-2265, 2010-2209, 2031-2311, 2042-2180, 2043-2196, 2067-2228, 2067-2322, 2135-2169

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
60/7512564CB1/ 1827	1-134, 1-170, 1-208, 1-211, 1-212, 1-213, 1-218, 1-236, 1-240, 1-252, 1-260, 1-261, 1-262, 1-264, 1-265, 1-268, 1-275, 1-280, 1-282, 1-283, 1-287, 1-288, 1-331, 1-390, 1-444, 1-447, 1-463, 1-478, 1-499, 1-1731, 2-261, 2-499, 3-277, 4-198, 4-210, 4-248, 4-252, 11-256, 13-301, 18-499, 23-201, 23-251, 23-264, 25-161, 25-292, 25-499, 26-266, 26-272, 26-298, 26-313, 27-189, 27-286, 28-266, 28-274, 28-336, 29-265, 29-305, 29-307, 29-347, 29-499, 30-217, 30-260, 30-274, 30-382, 32-217, 32-279, 32-311, 32-499, 33-143, 33-297, 35-134, 35-277, 35-283, 35-288, 35-322, 36-280, 37-174, 37-329, 38-303, 38-310, 38-317, 47-306, 50-294, 51-232, 51-309, 52-180, 53-357, 54-189, 54-325, 54-499, 58-338, 62-299, 64-338, 111-363, 116-220, 116-385, 116-418, 135-480, 164-404, 174-444, 210-487, 210-499, 212-408, 212-455, 214-468, 216-479, 223-484, 236-437, 236-452, 236-499, 241-485, 242-499, 252-443, 269-495, 269-499, 270-468, 275-770, 310-499, 353-499, 497-1089, 498-733, 498-736, 498-746, 498-750, 498-757, 498-910, 498-1008,
	498-1030, 498-1059, 498-1074, 498-1108, 498-1174, 500-720, 500-1042, 500-1049, 500-1059, 500-1102, 501-1059, 503-1058, 504-978, 506-1057, 507-610, 512-1169, 517-1027, 524-1058, 525-1063, 525-1080, 525-1113, 525-1169, 527-1009, 527-1063, 527-1091, 527-1093, 527-1100, 527-1117, 527-1126, 527-1129, 527-1144, 527-1149, 527-1167, 527-1182, 527-1203, 530-1053, 530-1109, 530-1136, 532-1130, 533-1195, 534-848, 536-1059, 554-1165, 554-1205, 558-706, 558-1188, 558-1215, 560-804, 560-1053, 560-1087, 562-1072, 565-1122, 568-1059, 571-937, 571-1059, 574-832, 574-1045, 574-1195, 576-846, 576-1063, 577-917, 579-1059, 580-1146, 584-864, 584-1095, 584-1167, 584-1260, 586-961, 588-788, 588-1059, 588-1100, 588-1133, 590-1138, 591-1233, 595-1059, 598-1212, 599-1115, 604-833, 605-1110, 607-1074, 611-1059, 613-798, 618-911, 618-1184, 620-917, 621-818, 621-881, 621-1146, 623-882, 624-1059, 625-874, 633-912, 634-1210, 635-910, 635-913, 635-1057, 635-1059, 635-1103, 636-1058, 651-1330, 653-1030, 653-1089, 653-1243, 654-1237,
	655-966, 657-1209, 662-782, 662-897, 665-1242, 668-784, 669-947, 673-775, 674-915, 674-931, 676-845, 678-968, 682-1069, 695-1374, 696-1015, 697-1063, 701-900, 701-991, 701-1290, 706-1374, 714-1376, 716-998, 718-1203, 721-849, 721-1059, 721-1234, 721-1294, 727-1351, 727-1375, 728-1259, 737-844, 743-1374, 744-1374, 747-1058, 747-1059, 747-1103, 747-1138, 747-1181, 747-1193, 747-1233, 747-1234, 747-1294, 748-1209, 750-1371, 751-1025, 755-967, 756-1011, 756-1053, 756-1177, 756-1259, 758-1004, 759-1374, 760-1008, 765-1371, 766-1409, 767-1269, 768-1375, 771-1371, 772-1374, 774-1250, 774-1371, 775-946, 775-1044, 776-1374, 778-1374, 779-1374, 779-1375, 783-1375, 785-1374, 786-1375, 788-1374, 789-1374, 790-1293, 791-1294, 794-1077, 799-1077, 800-995, 800-1409, 804-952, 806-1084, 814-1070, 819-1513, 822-1071, 822-1463, 824-1121, 824-1374, 824-1424, 824-1471, 825-1375, 827-1033, 830-1375, 839-1480, 845-1474, 848-1090, 848-1471, 855-1101, 857-1374, 864-1046, 869-1145, 869-1462, 871-1250, 871-1422, 875-1091, 877-1375,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	890-1049, 893-1516, 894-1161, 900-1155, 900-1266, 900-1348, 904-1374, 911-1299, 913-1227, 918-1218, 918-1250, 919-1178, 919-1215, 919-1218, 919-1516, 921-1201, 921-1368, 921-1501, 923-1382, 928-1669, 931-1214, 932-1337, 935-1257, 936-1209, 937-1678, 944-1538, 946-1243, 947-1538, 948-1603, 948-1670, 949-1196, 949-1519, 953-1145, 953-1207, 958-1211, 958-1537, 960-1139, 960-1244, 960-1385, 960-1437, 960-1593, 961-1240, 961-1697, 964-1494, 975-1250, 975-1668, 975-1679, 978-1241, 978-1452, 983-1665, 990-1716, 994-1239, 996-1179, 998-1244, 1001-1194, 1002-1638, 1003-1589, 1004-1697, 1008-1620, 1014-1610, 1015-1283, 1015-1299, 1016-1618, 1024-1237, 1024-1619, 1024-1620, 1024-1697, 1026-1442, 1026-1464, 1026-1478, 1026-1513, 1029-1283, 1031-1538, 1032-1678, 1034-1179, 1034-1298, 1035-1246, 1035-1294, 1035-1298, 1035-1461, 1038-1257, 1038-1275, 1038-1605, 1038-1669, 1039-1538, 1039-1617, 1042-1315, 1043-1530, 1046-1316, 1049-1356, 1050-1217, 1050-1314, 1050-1357, 1055-1410, 1060-1294, 1060-1305, 1060-1400, 1061-1400,
	1063-1556, 1068-1537, 1069-1651, 1072-1530, 1078-1533, 1078-1689, 1078-1737, 1082-1632, 1082-1655, 1082-1670, 1083-1619, 1090-1340, 1093-1649, 1100-1617, 1101-1357, 1101-1569, 1101-1717, 1104-1193, 1104-1533, 1104-1545, 1108-1314, 1110-1700, 1112-1443, 1112-1452, 1113-1530, 1114-1697, 1117-1670, 1117-1720, 1119-1530, 1119-1596, 1123-1323, 1123-1607, 1124-1377, 1127-1637, 1128-1338, 1128-1720, 1129-1751, 1130-1578, 1134-1309, 1134-1666, 1135-1318, 1135-1426, 1135-1434, 1135-1487, 1135-1742, 1138-1551, 1138-1634, 1139-1206, 1141-1745, 1143-1618, 1144-1592, 1151-1552, 1151-1615, 1152-1257, 1154-1356, 1162-1669, 1162-1727, 1164-1393, 1164-1396, 1166-1400, 1169-1766, 1171-1618, 1174-1530, 1176-1499, 1176-1748, 1177-1616, 1188-1750, 1190-1738, 1195-1438, 1198-1400, 1198-1678, 1199-1459, 1199-1550, 1200-1474, 1206-1731, 1206-1735, 1213-1538, 1224-1442, 1225-1485, 1225-1503, 1227-1750, 1257-1750, 1260-1776, 1264-1507, 1264-1527, 1270-1534, 1271-1529, 1276-1738, 1279-1725, 1288-1741, 1291-1430, 1294-1527, 1297-1505,
	1298-1737, 1298-1741, 1300-1728, 1304-1723, 1307-1752, 1308-1599, 1310-1720, 1313-1597, 1313-1667, 1314-1585, 1315-1727, 1320-1621, 1320-1720, 1321-1691, 1322-1720, 1323-1720, 1324-1582, 1326-1586, 1326-1591, 1326-1605, 1326-1618, 1326-1625, 1326-1720, 1327-1516, 1327-1536, 1327-1720, 1328-1720, 1328-1742, 1329-1603, 1332-1720, 1333-1443, 1333-1526, 1333-1533, 1333-1535, 1333-1614, 1333-1615, 1333-1655, 1333-1661, 1333-1678, 1333-1688, 1333-1718, 1333-1720, 1333-1721, 1333-1722, 1333-1723, 1333-1724, 1333-1725, 1333-1726, 1333-1728, 1333-1729, 1334-1720, 1334-1723, 1336-1718, 1342-1499, 1342-1558, 1342-1613, 1342-1617, 1342-1632, 1342-1720, 1343-1588, 1345-1720, 1346-1720, 1347-1720, 1348-1564, 1348-1725, 1349-1727, 1350-1729, 1352-1604, 1352-1635, 1352-1720, 1353-1589, 1353-1643, 1356-1721, 1356-1750, 1356-1770, 1357-1608, 1358-1625, 1360-1728, 1360-1741, 1361-1725, 1361-1729, 1363-1728, 1366-1620, 1366-1720, 1370-1636, 1370-1721, 1373-1632, 1373-1720, 1374-1570, 1374-1617, 1374-1660, 1374-1681, 1374-1734,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1377-1722, 1378-1667, 1379-1720, 1381-1725, 1382-1647, 1382-1720, 1386-1716, 1386-1730, 1387-1720, 1389-1720, 1392-1721, 1392-1728, 1394-1658, 1396-1477, 1396-1750, 1400-1677, 1400-1729, 1401-1734, 1403-1750, 1404-1601, 1407-1533, 1407-1708, 1407-1723, 1409-1720, 1411-1668, 1411-1689, 1415-1571, 1418-1722, 1420-1689, 1420-1699, 1421-1676, 1423-1720, 1424-1729, 1429-1734, 1430-1628, 1436-1720, 1437-1720, 1442-1720, 1444-1720, 1444-1724, 1445-1720, 1446-1720, 1447-1720, 1448-1720, 1450-1720, 1451-1720, 1451-1730, 1454-1679, 1454-1725, 1454-1730, 1456-1720, 1461-1701, 1461-1720, 1466-1701, 1466-1735, 1466-1736, 1468-1723, 1473-1720, 1474-1735, 1475-1740, 1477-1720, 1478-1711, 1478-1720, 1479-1734, 1483-1720, 1486-1729, 1493-1720, 1494-1731, 1495-1720, 1498-1725, 1500-1720, 1500-1749, 1501-1729, 1514-1720, 1517-1720, 1518-1720, 1519-1720, 1520-1720, 1520-1726, 1521-1722, 1522-1724, 1525-1720, 1532-1601, 1532-1720, 1532-1742, 1540-1725, 1541-1748, 1544-1776, 1544-1803, 1552-1720, 1554-1666, 1554-1673, 1554-1719,
	1554-1733, 1555-1733, 1557-1720, 1557-1734, 1559-1776, 1562-1728, 1565-1713, 1565-1734, 1567-1736, 1577-1827, 1578-1731, 1579-1720, 1582-1724, 1583-1711, 1583-1727, 1587-1720, 1633-1720, 1647-1827, 1661-1720, 1666-1728, 1669-1776, 1693-1726
61/7512646CB1/ 714	1-217, 1-240, 1-255, 1-268, 1-275, 1-276, 1-292, 2-186, 2-244, 2-254, 2-262, 2-275, 2-292, 2-297, 2-389, 2-417, 2-575, 2-714, 3-248, 3-259, 3-573, 4-271, 4-445, 5-294, 5-442, 5-557, 6-232, 6-294, 7-230, 7-327, 8-200, 8-217, 8-233, 8-282, 8-534, 8-550, 9-256, 10-194, 10-225, 10-239, 10-254, 10-262, 10-267, 10-298, 10-301, 10-311, 10-521, 10-568, 13-183, 13-219, 13-253, 13-254, 13-257, 13-268, 13-273, 13-275, 13-281, 13-283, 13-287, 13-291, 13-340, 13-358, 13-416, 13-504, 13-568, 14-298, 15-236, 15-276, 15-316, 15-347, 16-251, 16-564, 17-234, 17-258, 17-263, 17-271, 17-275, 17-292, 17-575, 18-193, 18-198, 18-229, 18-258, 18-264, 18-275, 18-281, 18-283, 18-286, 18-289, 18-296, 18-301, 18-308, 18-313, 18-485, 18-521, 18-573, 19-222, 19-244, 19-248, 19-267, 19-281, 19-287, 19-302, 19-305, 19-317, 19-335, 19-532, 19-543, 20-284, 20-294, 20-305, 20-306, 20-500, 21-236, 21-306, 22-575, 23-240, 23-249, 23-270, 23-288, 23-344, 23-575, 24-262, 24-263, 24-283, 24-287, 24-297, 25-271, 25-337, 26-232, 26-242, 26-256,
	26-264, 26-273, 26-276, 26-280, 26-291, 26-300, 26-306, 26-309, 26-310, 26-312, 26-314, 26-338, 26-410, 26-426, 26-527, 27-285, 27-301, 27-312, 27-333, 27-575, 28-265, 28-274, 28-292, 28-319, 28-575, 29-171, 29-174, 29-206, 29-255, 29-258, 29-266, 29-268, 29-273, 29-278, 29-283, 29-285, 29-287, 29-289, 29-290, 29-292, 29-293, 29-301, 29-302, 29-303, 29-320, 29-323, 29-330, 29-335, 29-503, 29-534, 29-549, 29-575, 30-209, 30-275, 30-282, 30-299, 30-301, 30-307, 30-310, 30-315, 30-322, 30-332, 30-337, 30-349, 30-498, 30-557, 30-575, 31-317, 31-331, 31-557, 31-575, 32-262, 32-265, 32-274, 32-355, 32-356, 33-289, 33-292, 33-302, 33-322, 33-327, 33-439, 34-285, 34-286, 34-299, 34-318, 35-245, 35-700, 36-199, 36-216, 36-269, 36-308, 36-319, 36-322, 36-389, 36-545, 36-550, 37-262, 37-398, 38-315, 38-575, 39-311, 39-320, 39-338, 39-532, 40-219, 40-254, 40-296, 40-307, 40-310, 40-326, 40-355, 40-363, 40-381, 41-301, 41-503, 41-540, 42-284, 42-310, 42-322, 42-534, 43-284, 43-287, 43-311, 43-575, 44-175, 44-245, 44-305,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	44-309, 45-294, 45-339, 45-496, 46-240, 46-344, 47-311, 47-315, 47-546, 47-562, 48-175, 48-261, 49-309, 50-271, 50-575, 51-521, 52-337, 59-470, 66-347, 69-377, 72-298, 80-559, 81-372, 98-298, 99-309, 111-299, 120-403, 123-244, 129-416, 139-403, 147-575, 157-449, 162-406, 176-411, 177-459, 177-491, 178-440, 182-437, 182-487, 192-456, 193-436, 193-454, 193-456, 193-464, 193-468, 200-447, 206-487, 221-485, 223-517, 241-479, 258-525, 261-575, 266-541, 269-545, 269-553, 272-539, 276-526, 276-556, 280-467, 285-556, 290-547, 290-561, 308-550, 309-574, 317-517, 326-562, 327-575, 335-490, 335-575, 336-567, 337-558, 337-575, 338-575, 341-575, 344-562, 346-568, 346-575, 350-519, 352-541, 353-575, 354-575, 360-559, 365-575, 402-525, 430-547
62/7512700CB1/ 1034	1-284, 1-1034, 94-651, 96-945, 221-647, 612-841, 778-919, 778-980, 778-989, 778-1010, 778-1027, 781-1025
63/7512707CB1/ 1071	1-255, 1-293, 1-1071, 5-289, 5-582, 8-253, 11-376, 13-206, 13-247, 13-309, 14-269, 14-294, 23-256, 23-260, 23-283, 23-313, 23-332, 23-412, 23-434, 23-462, 23-506, 23-528, 23-536, 23-611, 29-298, 31-304, 31-624, 38-254, 38-278, 43-617, 43-665, 47-689, 48-338, 70-687, 144-650, 206-658, 266-436, 269-689, 286-766, 338-816, 343-594, 375-671, 389-585, 430-648, 434-989, 439-722, 465-711, 466-752, 490-762, 495-726, 506-754, 514-785, 556-820, 560-874, 614-918, 655-998, 671-899, 815-916, 993-1069
64/7512710CB1/ 1399	1-179, 1-302, 1-1399, 180-1032, 191-669, 196-447, 228-524, 242-438, 246-850, 271-846, 283-501, 292-575, 302-883, 306-886, 318-691, 319-605, 319-950, 333-903, 336-953, 343-615, 348-579, 359-607, 364-964, 367-638, 392-1008, 409-673, 413-727, 413-1026, 416-962, 447-1090, 461-1094, 462-1050, 467-935, 473-976, 476-1336, 478-1071, 478-1125, 482-1099, 483-1035, 491-830, 492-877, 501-1126, 506-1000, 514-1114, 516-1045, 521-1048, 524-752, 533-1177, 534-909, 537-1107, 540-1206, 545-944, 546-1007, 559-1112, 599-1346, 600-1119, 603-1010, 606-1054, 616-927, 635-1082, 648-1090, 653-922, 659-1119, 664-1225, 668-769, 670-885, 670-1332, 681-1127, 687-1327, 716-1281, 775-1337, 781-1086, 781-1176, 783-1078, 792-1046, 792-1059, 809-1015, 809-1027, 809-1033, 820-1041, 828-1055, 841-1061, 841-1157, 847-1351, 861-1096, 862-1361, 867-1385, 867-1399, 892-1130, 893-1192, 895-1399, 904-1369, 906-1326, 916-1072, 916-1396, 919-1398, 921-1396, 929-1396, 933-1126, 936-1170, 941-1249, 943-1399, 945-1399, 948-1399, 951-1329, 954-1399,
	956-1399, 961-1395, 965-1209, 967-1399, 970-1396, 977-1399, 978-1393, 978-1396, 981-1397, 982-1396, 990-1396, 1003-1289, 1006-1285, 1007-1271, 1010-1327, 1018-1396, 1019-1229, 1019-1399, 1021-1399, 1025-1398, 1026-1396, 1039-1399, 1041-1399, 1046-1322, 1051-1396, 1053-1393, 1066-1168, 1082-1396, 1090-1399, 1094-1399, 1100-1399, 1102-1399, 1122-1399, 1126-1396, 1140-1396, 1142-1383, 1162-1399, 1163-1399, 1172-1399, 1215-1398, 1332-1399
65/7512884CB1/ 1634	1-245, 1-1634, 296-559, 410-648

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
66/7512931CB1/ 1631	1-397, 1-419, 1-542, 17-207, 17-383, 17-1631, 29-307, 30-295, 30-313, 47-316, 58-347, 79-307, 109-306, 178-383, 188-425, 213-510, 213-542, 254-542, 304-546, 313-546, 424-1274, 559-1089, 564-800, 564-822, 602-898, 616-927, 617-896, 618-1040, 628-1088, 646-1230, 652-1088, 654-795, 691-953, 697-1254, 709-964, 711-866, 714-1621, 723-936, 739-1009, 746-1600, 746-1621, 755-1621, 782-1064, 788-1621, 789-1488, 801-1287, 809-1407, 819-1454, 844-1285, 850-1128, 850-1407, 871-1576, 885-1620, 886-1433, 908-1450, 908-1580, 918-1621, 923-1514, 941-1180, 942-1491, 946-1321, 952-1203, 953-1621, 989-1256, 999-1234, 999-1249, 1012-1133, 1052-1317, 1054-1484, 1069-1318, 1074-1631, 1088-1538, 1089-1575, 1094-1319, 1100-1359, 1122-1621, 1128-1402, 1132-1332, 1132-1341, 1138-1563, 1139-1369, 1151-1631, 1156-1621, 1164-1618, 1165-1621, 1172-1613, 1188-1612, 1198-1631, 1202-1621, 1206-1621, 1219-1621, 1224-1631, 1266-1563, 1287-1545, 1291-1442, 1301-1611, 1304-1631, 1309-1621, 1317-1621, 1318-1621, 1319-1621, 1320-1628, 1339-1565,
	1355-1629, 1363-1617, 1367-1619, 1402-1621, 1412-1621, 1423-1621, 1431-1631, 1441-1631, 1442-1631, 1485-1621, 1508-1631, 1526-1631, 1556-1621
67/7512933CB1/ 808	1-284, 1-397, 1-419, 1-507, 1-525, 1-528, 1-531, 1-536, 1-540, 1-545, 1-570, 1-612, 1-632, 3-291, 17-207, 17-237, 17-285, 17-292, 17-314, 17-383, 17-808, 25-727, 29-205, 29-307, 30-295, 30-313, 32-295, 47-316, 58-347, 79-307, 109-306, 138-264, 167-579, 171-580, 177-796, 178-383, 188-425, 213-426, 254-426, 334-612, 339-788, 349-497, 362-618, 378-806, 390-496, 401-808, 425-793, 425-808, 426-808, 447-744, 468-726, 472-623, 482-792, 485-802, 490-802, 498-802, 499-802, 500-802, 501-808, 520-746, 536-808, 544-798, 548-808, 583-802, 593-802, 604-802, 612-808, 622-808, 623-808, 666-802, 689-808, 707-808, 737-806
68/7512942CB1/ 657	1-191, 1-221, 1-268, 1-269, 1-271, 1-275, 1-276, 1-657, 13-189, 13-276, 14-276, 16-276, 31-275, 63-276, 93-276, 122-248, 187-652, 206-652, 264-586, 273-657, 274-642, 274-657, 275-657, 296-593, 317-575, 321-472, 331-641, 334-651, 339-651, 347-651, 348-651, 349-651, 350-657, 369-595, 385-657, 393-647, 397-657, 432-651, 442-651, 453-651, 461-657, 471-657, 472-657, 515-651, 538-657, 556-657, 586-655
69/7513736CB1/ 1158	1-654, 298-1158

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
70/7512779CB1/ 2125	1-116, 1-122, 1-131, 1-132, 1-148, 1-149, 1-159, 1-167, 1-188, 1-197, 1-204, 1-224, 1-242, 1-254, 1-261, 1-274, 1-282, 1-319, 1-351, 1-358, 1-364, 1-376, 1-377, 1-384, 1-395, 1-404, 1-407, 1-421, 1-431, 1-433, 1-435, 1-439, 1-447, 1-449, 1-452, 1-456, 1-460, 1-464, 1-469, 1-477, 1-481, 1-482, 1-483, 1-485, 1-488, 1-489, 1-494, 1-496, 1-498, 1-500, 1-501, 1-503, 1-510, 1-531, 1-534, 1-536, 1-542, 1-548, 1-584, 1-602, 1-613, 2-272, 4-499, 4-628, 6-544, 14-1798, 22-229, 55-656, 80-638, 88-672, 90-682, 91-641, 114-711, 115-398, 120-278, 126-629, 136-677, 144-683, 151-802, 159-376, 162-702, 163-696, 164-594, 177-704, 177-729, 200-727, 213-427, 223-487, 229-708, 234-570, 237-399, 239-417, 239-781, 252-761, 267-602, 271-524, 271-703, 272-417, 280-939, 287-789, 288-709, 290-743, 314-834, 318-844, 318-902, 320-867, 329-908, 332-722, 339-722, 348-883, 350-897, 360-541, 360-911, 366-569, 375-553, 375-554, 395-750, 398-985, 405-997, 406-679, 406-680, 406-936, 409-1013, 410-951, 412-1035, 416-877, 420-1048, 424-629,
	424-797, 424-935, 426-974, 432-631, 436-723, 440-942, 442-584, 444-983, 448-722, 450-654, 451-964, 457-743, 466-722, 467-998, 474-628, 474-702, 478-702, 478-739, 479-1137, 480-1029, 484-667, 485-1026, 493-908, 493-1061, 498-1042, 515-639, 515-765, 515-1041, 515-1046, 517-697, 517-1048, 526-769, 527-994, 527-1030, 528-1185, 533-1048, 539-1188, 544-704, 545-797, 546-839, 548-1139, 551-1081, 558-1030, 562-844, 563-1163, 565-676, 565-744, 567-1134, 571-1097, 577-1231, 581-840, 581-1185, 595-809, 595-936, 596-1021, 600-1134, 601-1046, 603-697, 606-789, 608-1152, 609-1050, 615-1160, 619-1228, 624-1158, 628-1276, 629-911, 633-1283, 639-720, 639-1191, 645-882, 651-1194, 653-846, 653-1181, 658-944, 658-1181, 658-1224, 663-1050, 663-1191, 664-995, 668-1173, 670-1211, 673-845, 676-919, 689-916, 699-1241, 699-1250, 699-1292, 702-849, 708-867, 709-1246, 709-1251, 712-1031, 716-1248, 729-1078, 732-976, 741-1260, 741-1292, 744-913, 752-1292, 766-1177, 766-1292, 768-994, 768-1086, 768-1287, 771-955, 775-1095, 777-1292,
	787-1292, 789-994, 794-1292, 795-1285, 799-910, 799-1052, 808-947, 809-913, 823-1047, 825-1172, 826-1191, 826-1292, 829-1161, 831-977, 833-1149, 837-1105, 847-1124, 847-1206, 847-1292, 852-1255, 853-1283, 867-1290, 867-1292, 868-977, 883-1091, 883-1211, 890-1199, 892-1008, 894-1103, 902-1292, 915-1078, 916-1185, 927-1220, 930-1177, 935-1292, 939-1292, 948-1058, 953-1220, 957-1211, 965-1133, 966-1292, 1010-1159, 1014-1177, 1017-1252, 1022-1196, 1029-1194, 1030-1292, 1035-1290, 1035-1291, 1039-1270, 1049-1224, 1051-1250, 1055-1220, 1073-1288, 1080-1288, 1087-1201, 1087-1202, 1099-1292, 1102-1290, 1124-1292, 1138-1268, 1178-1292, 1247-1516, 1287-1397, 1287-1652, 1287-1686, 1287-1778, 1287-1794, 1288-1868, 1291-1945, 1297-1796, 1299-1656, 1311-1913, 1320-1903, 1321-1600, 1322-1753, 1334-1763, 1334-1917, 1342-1871, 1356-1576, 1356-1672, 1356-1793, 1356-1879, 1356-1953, 1357-1853, 1361-1893, 1363-1826, 1364-1903, 1366-1575, 1366-1589, 1368-1586, 1368-1875, 1372-1843, 1373-1881, 1374-1656, 1377-1916, 1381-1734,
	1381-1921, 1383-1586, 1385-1918, 1391-1689, 1391-1933, 1395-1596, 1398-1566, 1418-1586, 1419-1617, 1435-1787, 1437-2050, 1450-1603, 1465-1630, 1468-1695, 1471-2125, 1483-1659, 1485-1665, 1486-1886, 1493-1903, 1517-1819, 1518-1688, 1525-1787, 1528-1757, 1530-1804, 1531-1707, 1532-1745

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
71/7512877CB1/ 1913	1-296, 4-331, 4-441, 4-512, 4-1905, 153-473, 211-1126, 230-528, 239-688, 436-650, 453-700, 477-668, 518-685, 721-1237, 802-1331, 805-1239, 823-1755, 834-1755, 841-1232, 843-1282, 853-1469, 862-1361, 868-1317, 872-1755, 883-1284, 897-1600, 902-1750, 905-1755, 909-1499, 909-1755, 912-1755, 913-1755, 914-1755, 917-1755, 923-1755, 925-1257, 929-1755, 930-1601, 935-1616, 952-1570, 979-1755, 986-1670, 989-1275, 1001-1236, 1005-1755, 1006-1755, 1007-1410, 1007-1605, 1007-1611, 1008-1362, 1023-1755, 1026-1755, 1029-1293, 1036-1663, 1043-1910, 1065-1344, 1065-1571, 1076-1339, 1077-1480, 1079-1333, 1079-1343, 1080-1303, 1080-1320, 1091-1473, 1091-1537, 1101-1350, 1103-1705, 1104-1663, 1105-1360, 1112-1423, 1117-1191, 1121-1758, 1127-1769, 1137-1423, 1138-1787, 1147-1842, 1149-1842, 1169-1449, 1169-1679, 1179-1832, 1185-1913, 1190-1571, 1202-1474, 1210-1840, 1220-1876, 1236-1437, 1240-1754, 1245-1785, 1246-1450, 1246-1480, 1246-1694, 1246-1702, 1246-1741, 1246-1781, 1249-1681, 1250-1743, 1254-1695, 1255-1522,
	1274-1559, 1283-1831, 1292-1860, 1295-1907, 1298-1557, 1298-1625, 1310-1637, 1310-1763, 1313-1913, 1315-1535, 1316-1909, 1344-1812, 1346-1887, 1360-1809, 1361-1672, 1362-1811, 1368-1617, 1368-1619, 1368-1661, 1368-1872, 1368-1913, 1370-1641, 1385-1850, 1385-1852, 1396-1871, 1403-1858, 1412-1911, 1414-1866, 1431-1913, 1432-1840, 1434-1583, 1434-1675, 1438-1683, 1440-1840, 1440-1855, 1449-1870, 1451-1906, 1454-1713, 1454-1723, 1454-1785, 1458-1911, 1461-1913, 1462-1807, 1464-1906, 1464-1913, 1471-1913, 1473-1815, 1478-1703, 1478-1716, 1478-1913, 1482-1877, 1483-1849, 1486-1907, 1494-1909, 1496-1913, 1498-1752, 1498-1913, 1501-1913, 1502-1906, 1504-1910, 1509-1911, 1510-1913, 1524-1773, 1526-1791, 1528-1794, 1533-1911, 1535-1904, 1542-1913, 1544-1911, 1544-1913, 1548-1790, 1548-1906, 1555-1829, 1558-1911, 1559-1911, 1561-1910, 1562-1910, 1564-1888, 1566-1910, 1566-1911, 1566-1913, 1571-1824, 1571-1905, 1571-1912, 1578-1911, 1585-1900, 1589-1905, 1589-1912, 1591-1913, 1601-1913, 1602-1911, 1613-1876, 1613-1912,
	1614-1883, 1621-1900, 1636-1840, 1641-1905, 1664-1812, 1687-1911, 1703-1905, 1721-1884, 1752-1913, 1801-1913, 1823-1902
72/7512782CB1/ 1183	1-232, 1-517, 3-200, 4-225, 5-203, 16-475, 26-227, 33-654, 34-338, 37-154, 37-248, 37-251, 148-243, 221-809, 308-750, 308-779, 318-757, 321-569, 340-904, 342-599, 342-600, 342-606, 342-609, 342-615, 342-743, 342-758, 342-816, 342-833, 342-894, 347-880, 352-620, 354-619, 358-597, 358-617, 358-796, 361-649, 401-826, 401-881, 408-896, 408-898, 410-796, 410-821, 422-666, 445-874, 451-796, 453-909, 464-909, 465-736, 472-796, 474-796, 484-796, 500-796, 504-796, 544-898, 548-772, 572-761, 572-806, 572-816, 572-894, 585-796, 586-796, 586-894, 606-796, 606-809, 606-925, 606-1183, 616-875, 622-796, 622-855, 627-796, 708-790, 730-796
73/7512784CB1/ 1074	1-201, 5-201, 7-201, 21-638, 39-156, 200-641, 200-670, 209-648, 212-460, 231-795, 233-490, 233-491, 233-497, 233-500, 233-506, 233-634, 233-649, 233-707, 233-724, 233-785, 238-771, 243-511, 245-510, 249-488, 249-508, 249-687, 252-540, 292-717, 292-772, 299-787, 299-789, 301-687, 301-712, 313-557, 336-765, 342-687, 344-800, 355-800, 356-627, 363-687, 365-687, 375-687, 391-687, 395-687, 435-789, 439-663, 463-652, 463-697, 463-707, 463-785, 476-687, 477-687, 477-785, 497-687, 497-700, 497-816, 497-1074, 507-766, 513-687, 513-746, 518-687, 599-681, 621-687

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
74/7512794CB1/ 1021	1-153, 1-236, 1-251, 1-252, 1-257, 1-259, 1-265, 1-267, 1-288, 1-302, 1-320, 1-1014, 2-254, 2-258, 2-267, 3-90, 3-132, 3-175, 3-185, 3-238, 3-240, 3-280, 3-298, 4-256, 4-292, 4-308, 5-188, 5-214, 5-253, 7-253, 12-208, 12-255, 12-280, 12-303, 13-133, 13-259, 16-219, 16-263, 16-302, 16-578, 39-322, 44-164, 48-292, 57-315, 193-338, 211-417, 409-793, 409-958, 410-1021, 416-907, 423-973, 424-672, 430-1013, 436-804, 441-989, 446-1013, 471-737, 476-686, 476-913, 495-789, 505-768, 505-769, 511-672, 517-786, 526-654, 526-1002, 534-1003, 546-774, 568-860, 570-850, 570-882, 586-1003, 596-914, 614-864, 617-784, 617-806, 617-843, 621-773, 636-822, 636-1014, 645-1003, 649-992, 650-825, 662-995, 690-949, 692-828, 707-944, 714-998, 725-999, 733-953, 740-767, 750-995, 757-998, 758-892, 758-971, 758-1021, 773-1018, 829-1021
75/7512886CB1/ 1127	1-230, 1-269, 1-1127, 30-395, 35-269, 35-270, 36-269, 36-306, 38-269, 38-281, 41-262, 42-315, 42-438, 42-600, 45-335, 45-447, 45-497, 46-156, 46-292, 46-492, 49-319, 49-368, 60-177, 87-530, 92-527, 436-855, 527-763, 527-787, 527-894, 527-939, 527-959, 527-1120, 529-1120, 563-775, 572-662, 572-773, 572-847, 606-1120, 620-894, 657-1120, 677-1127, 682-1073, 683-1045, 710-1045, 711-1119, 771-1120, 797-1110, 853-1110, 855-1127, 878-1071, 878-1120, 897-1114, 908-1117, 909-1110, 974-1120, 977-1118, 1005-1041
76/7512929CB1/ 1855	1-247, 1-281, 1-316, 1-1855, 3-265, 7-752, 10-463, 45-186, 45-242, 186-325, 325-523, 325-938, 330-612, 341-522, 358-588, 358-1163, 369-878, 378-1079, 386-903, 387-887, 390-866, 393-670, 396-997, 398-624, 417-936, 425-964, 435-1095, 436-687, 465-926, 465-1112, 486-738, 487-758, 488-760, 505-917, 510-927, 517-647, 522-759, 544-664, 550-777, 550-1061, 587-987, 600-1192, 603-1175, 612-946, 622-943, 639-1056, 646-1317, 657-1247, 658-879, 661-936, 661-989, 671-1266, 674-1266, 679-1168, 679-1242, 690-1078, 695-972, 696-977, 696-1345, 698-1274, 707-1126, 738-1003, 744-996, 746-1029, 747-1032, 748-869, 768-963, 774-1310, 779-1394, 785-1036, 792-1419, 793-1231, 793-1270, 800-1334, 824-1076, 844-1409, 855-1049, 857-1176, 863-1078, 872-1094, 876-1400, 883-1391, 889-1419, 890-1447, 901-1231, 902-1134, 905-1365, 905-1476, 912-1337, 914-1210, 926-1476, 929-1207, 933-1211, 933-1219, 937-1434, 943-1556, 948-1464, 948-1529, 949-1392, 949-1575, 954-1439, 954-1480, 963-1220, 970-1476, 970-1534, 974-1232, 976-1846, 994-1553,
	995-1270, 1001-1231, 1002-1570, 1011-1603, 1030-1601, 1038-1317, 1040-1553, 1044-1503, 1045-1285, 1053-1491, 1058-1487, 1067-1321, 1069-1289, 1069-1329, 1082-1321, 1082-1443, 1088-1534, 1091-1736, 1105-1371, 1106-1382, 1107-1347, 1108-1381, 1127-1376, 1145-1553, 1149-1374, 1152-1477, 1155-1736, 1156-1662, 1163-1770, 1171-1669, 1171-1811, 1172-1415, 1178-1438, 1178-1719, 1179-1330, 1180-1444, 1193-1553, 1207-1797, 1216-1445, 1231-1623, 1235-1855, 1236-1510, 1237-1484, 1237-1635, 1237-1774, 1238-1539, 1262-1542, 1266-1498, 1271-1537, 1273-1542, 1274-1796, 1285-1563, 1288-1539, 1288-1855, 1297-1743, 1320-1542, 1323-1797, 1325-1588, 1329-1732, 1332-1855, 1334-1575, 1344-1840, 1345-1855, 1349-1842, 1358-1744, 1363-1543, 1364-1661, 1367-1855, 1371-1855, 1384-1848, 1387-1658, 1388-1843, 1389-1850, 1400-1784, 1404-1846, 1405-1767, 1408-1846, 1410-1846, 1411-1845, 1415-1855, 1417-1855, 1425-1692, 1428-1842, 1428-1853, 1429-1846, 1431-1842, 1432-1623, 1432-1624, 1434-1842, 1434-1855, 1436-1851, 1437-1822, 1438-1843,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1440-1817, 1442-1732, 1442-1843, 1443-1842, 1445-1842, 1446-1843, 1451-1768, 1451-1841, 1451-1855, 1453-1541, 1453-1678, 1453-1721, 1453-1785, 1453-1855, 1454-1709, 1454-1846, 1461-1711, 1466-1672, 1476-1544, 1478-1744, 1491-1843, 1493-1843, 1494-1846, 1495-1842, 1503-1747, 1513-1719, 1519-1695, 1523-1842, 1548-1778, 1548-1787, 1548-1846, 1548-1855, 1550-1848, 1569-1842, 1572-1842, 1635-1846, 1646-1845, 1647-1855, 1650-1842, 1651-1845, 1652-1842, 1679-1842, 1692-1842, 1697-1846, 1709-1855

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
39	7511289CB1	OVARNOT03
40	7511056CB1	THYRNOT03
41	7511567CB1	LUNGNON07
42	7511651CB1	ADRETUT01
44	7512181CB1	LUNGNOT12
45	7511726CB1	FTUBTUT02
46	7511057CB1	PLACNOT02
47	7511078CB1	COLNTUT03
48	7511505CB1	EPIPNON05
49	7511552CB1	LIVRTUE01
50	7511722CB1	PROSNOT16
51	7511489CB1	THP1NOB01
52	7511497CB1	PROSTMT05
53	7511498CB1	HNT2RAT01
54	7511612CB1	MENITUT03
55	7511624CB1	PHOSDNV10
56	7511626CB1	PHOSDNV10
57	7512885CB1	BRSTNOT02
58	7511965CB1	GBLADIT02
59	7512403CB1	PHOSDNV44
60	7512564CB1	PROTDNV21
61	7512646CB1	OVARTUT04
62	7512700CB1	LIVRBCT01
63	7512707CB1	ISLTNOT01
64	7512710CB1	BRAINON01
65	7512884CB1	BRAINOT03
66	7512931CB1	BRAZNOT01
67	7512933CB1	THYRNOT10
68	7512942CB1	SININOT04
70	7512779CB1	BRSTTUT22
71	7512877CB1	MYOMNOT01
72	7512782CB1	FIBPFEN06
73	7512784CB1	FIBPFEN06
74	7512794CB1	SCOMDIT01
75	7512886CB1	SINTBST01
76	7512929CB1	LUNGNOT12

Table 6

Library	Vector	Library Description
ADRETUT01	PSPORT	Library was constructed using RNA isolated from right adrenal tumor tissue removed from a 50-year-old Turkish male during aunilateral adrenalectomy. Pathology indicated a metastatic renal cell carcinoma that formed a circumscribed, spongy, hemorrhagic nodule situated in the region of the medulla. The patient presented with corticoadrenal insufficiency, incisional hernia,and non-alcoholic steato hepatitis. Patient history included renal cell carcinoma. Family history included liver cancer.
BRAINON01	PSPORT1	Library was constructed and normalized from 4.88 million independent clones from a brain tissue library. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228, except that a significantly longer (48-hour) reannealing hybridization was used.
BRAINOT03	PSPORT1	Library was constructed using RNA isolated from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRAZNOT01	pINCY	Library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from a 45-year-old Caucasian female who died from a dissecting aortic aneurysm and ischemic bowel disease. Pathology indicated mild arteriosclerosis involving the cerebral cortical white matter and basal ganglia. Grossly, there was mild meningeal fibrosis and mild focal atherosclerotic plaque in the middle cerebral artery, as well as vertebral arteries bilaterally. Microscopically, the cerebral hemispheres, brain stem and cerebellum reveal focal areas in the white matter that contain blood vessels that were barrel-shaped, hyalinized, with hemosiderin-laden macrophages in the Virchow-Robin space. In addition, there were scattered neurofibrillary tangles within the basolateral nuclei of the amygdala. Patient history included mild atheromatosis of aorta and coronary arteries, bowel and liver infarct due to aneurysm, physiologic fatty liver associated with obesity, mild diffuse emphysema, thrombosis of
		mesenteric and portal veins, cardiomegaly due to hypertrophy of left ventricle, arterial hypertension, acute pulmonary edema, splenomegaly, obesity (300 lb.), leiomyoma of uterus, sleep apnea, and iron deficiency anemia.
BRSTNOT02	PSPORT1	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocysytic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.

Table 6

Library	Vector	Library Description
BRSTTUT22	pINCY	Library was constructed using RNA isolated from right breast tumor tissue removed from a 59-year-old Caucasian female during unilateral extended simple mastectomy. Pathology indicated invasive lobular carcinoma with extension into ducts forming an ill-defined mass situated in the right breast biopsy cavity site. The non-neoplastic breast parenchyma displays papillomatosis. Prior right breast biopsy indicated invasive grade 3, nuclear grade 3, invasive and in situ ductal carcinoma. Estrogen and progesterone immunostains were positive in the neoplastic cells. Patient history included cirrhosis of the liver, esophageal ulcer, hyperlipidemia, and neuropathy caused by an unspecified disease. Previous surgeries included segmental lung resection and a liver transplant. Patient medications included Prograf, prednisone, Imuran, Lozol, Zantac, Estraderm patches, and Provera.
COLNTUT03	pINCY	Library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma. One lymph node contained metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer, and prostate cancer.
EPIPNON05	pINCY	This normalized prostate epithelial cell tissue library was constructed from 2.36 million independent clones from a prostate epithelial cell tissue library. Starting RNA was made from untreated prostatic epithelial cell issue removed from a 17-year-old Hispanic male. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used.
FIBPFEN06	pINCY	The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a prostate fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round)reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation.

Table 6

Library	Vector	Library Description
FTUBTUT02	pINCY	Library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma was present a the cul-de-sac tumor. Patient history included medullary carcinoma of the thyroid and myocardial infarction.
GBLADIT02	pINCY	The library was constructed using RNA isolated from diseased gallbladder tissue removed from a 18-year-old Caucasian female during cholecystectomy and incidental appendectomy. Pathology indicated acute and chronic cholecystitis with cholelithiasis. The gallbladder contained multiple fragments of stony material. The appendix showed lymphoid hyperplasia. The patient presented with abdominal pain, nausea, and vomiting. Patient history included Chlamydia, extrinsicasthma, and cesarean delivery (x3). Family history included benign hypertension, acute myocardial infarction, and atherosclerotic coronary artery disease.
HNT2RAT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
LIVRBCT01	PBLUESCRIPT	Library was constructed using RNA isolated from the liver tissue of a patient with primary biliary cirrhosis who had a liver transplant.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.

Table 6

Library	Vector	Library Description
LUNGNON07	pINCY	This normalized lung tissue library was constructed from 5.1 million independent clones from a lung tissue library. Starting RNA was made from RNA isolated from lung tissue. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LUNGNOT12	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 78-year-old Caucasian male during a segmental lung resection and regional lymph node resection. Pathology indicated fibrosis pleura was puckered, but not invaded. Pathology for the associated tumor tissue indicated an invasive pulmonary grade 3 adenocarcinoma. Patient history included cerebrovascular disease, arteriosclerotic coronary artery disease, thrombophlebitis, chronic obstructive pulmonary disease, and asthma. Family history included intracranial hematoma, cerebrovascular disease, arteriosclerotic coronary artery disease, and type I diabetes.
MENITUT03	pINCY	Library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer .
MYOMNOT01	PSPORT1	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 43-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Family history included lung cancer, stroke, type II diabetes, hepatic lesion, chronic liver disease, hyperlipidemia, congenital heart anomaly, and mitral valve prolapse.
OVARNOT03	PSPORT1	Library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
OVARTUT04	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 53-year-old Caucasian female during a total abdominal hysterectomy, removal of the fallopian tubes and ovaries, regional lymph node excision, peritoneal tissue destruction, and incidental appendectomy. Pathology indicated grade 1 transitional cell carcinoma of the right ovary. The left ovary had a hemorrhagic corpus luteum. The uterus had multiple leiomyomas (1 submucosal, 11 intramural), and the endometrium was inactive. The cul-de-sac contained abundant histiocytes and rare clusters of mesothelial cells. Patient history included breast fibrosclerosis and chronic stomach ulcer. Family history included acute stomach ulcer with perforation, breast cancer, bladder cancer, rectal/anal cancer, benign hypertension, coronary angioplasty, and hyperlipidemia.

Table 6

Library	Vector	Library Description
PHOSDNV10	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.
PHOSDNV44	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.
PLACNOT02	pINCY	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
PROSTMT05	pINCY	The library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during a radical prostatectomy, regional lymph node excision, and prostate needle biopsy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason grade 5+4, forming a predominant mass involving the left side peripherally with extension into the right posterior superior region.The tumor invaded and perforated the capsule to involve periprostatic tissue in theleft posterior superior region. The left inferior and superior posterior surgical margins were positive. One (of 9) left pelvic lymph nodes was metastatically involved. The patient presented with elevated prostate specific antigen (PSA). Patient history included calculus of the kidney. Family history included breast cancer and lung cancer.
PROTDNV21	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled small intestine tissue removed from a Caucasian male fetus (donor A) who died at 23 weeks' gestation from premature birth; from lung tissue removed from a Caucasian male fetus (donor B) who died from fetal demise; from pleura tumor tissue removed from a 55-year-old Caucasian female (donor C) during a complete pneumonectomy; from frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor D) during excision of cerebral meningeal lesion; from liver tumor tissue removed from a 72-year-old Caucasian male (donor E) during partial hepatectomy; from pooled fetal brain tissue removed from a Caucasian male fetus (donor F) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor G), who died at 23 weeks' gestation from premature birth; from pooled fetal kidney tissue removed from 59, 20-33-week-old
		male and female fetuses who died from spontaneous abortion; from pooled thymus tissue removed from 9, 18-32-year-old male and female donors who died from sudden death; and from pooled fetal liver tissue removed from 32, 18-24-week-old male and female fetuses. For donor A, serologies were negative. Family history included diabetes in the mother. For donor B, Serologies were negative. For donor C, pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bosellated mass replacing the right lower lobe and a portion of the middle lobe. Multiple nodules comprising the tumor show near total necrosis. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. The patient presented with shortness of breath. Patient history included peptic ulcer disease, normal delivery, anemia, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, hemorrhoidectomy, endoscopic

Table 6

Library	Vector	Library Description
		excision of lung lesion, and appendectomy. Patient medications included Megace, tamoxifen, and Pepcid. Family history included multiple sclerosis in the mother; atherosclerotic coronary artery disease and type II diabetes in the father; and breast cancer in the grandparent(s). For donor D, pathology indicated neuroectodermal tumor with advanced ganglionic differentiation. The lesion was only moderately cellular but was mitotically active with a high MIB-I labelling index. Neuronal differentiation was widespread and advanced. Multinucleate and dysplastic-appearing forms were readily seen. The glial element was less prominent. The patient presented with motor seizures. Family history included hypertension in the grandparent(s). For donor E, pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and
		tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father. For donor F and G, Serologies were negative for both donors and family history for donor G included diabetes in the mother.
SCOMDIT01	pINCY	Library was constructed using RNA isolated from diseased spinal cord tissue removed from the base of the medulla of a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema.
SININOT04	pINCY	Library was constructed using RNA isolated from diseased ileum tissue obtained from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included enteritis of the small intestine.
SINTBST01	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from an 18-year-old Caucasian female during bowel anastomosis. Pathology indicated Crohn's disease of the ileum, involving 15 cm of the small bowel. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
THP1NOB01	PBLUESCRIPT	Library was constructed using RNA isolated from cultured, unstimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).

Table 6

Library	Vector	Library Description
THYRNOT03	pINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
THYRNOT10	pINCY	Library was constructed using RNA isolated from the diseased left thyroid tissue removed from a 30-year-old Caucasian female during a unilateral thyroid lobectomy and parathyroid reimplantation. Pathology indicated lymphocytic thyroiditis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
39	7511289	1260704H1	SNP00039681	10	1470	C	C	A	noncoding	n/d	n/d	n/d	n/d
39	7511289	1391688H1	SNP00115026	178	667	T	T	C	P202	n/a	n/a	n/a	n/a
39	7511289	1399006H1	SNP00111942	202	376	C	C	T	G105	n/d	n/a	n/a	n/a
39	7511289	1849722T6	SNP00039681	258	1471	C	C	A	noncoding	n/d	n/d	n/d	n/d
39	7511289	2132554R6	SNP00039681	10	1476	C	C	A	noncoding	n/d	n/d	n/d	n/d
39	7511289	2829206T6	SNP00039681	241	1472	C	C	A	noncoding	n/d	n/d	n/d	n/d
39	7511289	3412913H1	SNP00060558	147	915	A	G	A	noncoding	0.99	0.98	0.93	0.92
39	7511289	7423309T1	SNP00111942	415	377	C	C	T	L106	n/d	n/a	n/a	n/a
39	7511289	8038111H1	SNP00060558	156	916	G	G	A	noncoding	0.99	0.98	0.93	0.92
40	7511056	1443673F6	SNP00124167	13	13	C	C	T	noncoding	n/a	n/a	n/a	n/a
40	7511056	1443673R1	SNP00061413	26	904	C	C	T	noncoding	n/a	n/a	n/a	n/a
41	7511567	1721376H1	SNP00009248	98	1454	C	T	C	G468	n/a	n/a	n/a	n/a
41	7511567	1721376H1	SNP00057378	158	1514	C	C	T	L488	0.36	0.49	0.31	0.31
41	7511567	1809703F6	SNP00037263	195	1085	T	T	C	D345	n/a	n/a	n/a	n/a
41	7511567	1919212H1	SNP00011634	95	1514	T	T	C	L488	n/d	n/a	n/a	n/a
41	7511567	2372004H1	SNP00068404	130	1613	C	C	T	D521	n/a	n/a	n/a	n/a
41	7511567	7385231H1	SNP00011634	57	1515	T	T	C	Y489	n/d	n/a	n/a	n/a
41	7511567	7761581J1	SNP00037263	143	1089	C	T	C	L347	n/a	n/a	n/a	n/a
42	7511651	1672333H1	SNP00009036	30	76	A	G	A	noncoding	0.36	n/a	n/a	n/a
42	7511651	3448924H1	SNP00036880	24	42	C	C	A	noncoding	0.8	n/a	n/a	n/a
44	7512181	2377009H1	SNP00044235	83	318	G	G	A	A99	n/a	n/a	n/a	n/a
44	7512181	2851653H1	SNP00044236	134	557	C	C	A	A179	0.97	0.9	0.96	0.93
45	7511726	1225412H1	SNP00132159	53	409	T	T	C	L123	n/a	n/a	n/a	n/a
45	7511726	1225412H1	SNP00132160	158	514	C	C	G	A158	n/a	n/a	n/a	n/a
45	7511726	1282158F6	SNP00003186	27	265	T	T	C	I75	n/a	n/a	n/a	n/a
45	7511726	1321651F6	SNP00132158	28	66	T	T	C	S9	n/a	n/a	n/a	n/a
45	7511726	1326847F6	SNP00092454	382	1081	C	C	T	noncoding	n/d	n/a	n/a	n/a
45	7511726	1444142H1	SNP00003185	35	73	C	C	T	A11	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
45	7511726	1927771T6	SNP00003187	38	1075	T	T	C	noncoding	n/a	n/a	n/a	n/a
45	7511726	3074414T6	SNP00092454	78	1077	C	C	T	noncoding	n/d	n/a	n/a	n/a
45	7511726	353323T6	SNP00092454	24	1097	C	C	T	noncoding	n/d	n/a	n/a	n/a
45	7511726	3781813H1	SNP00147540	238	277	A	A	G	D79	n/a	n/a	n/a	n/a
45	7511726	4549457T1	SNP00092454	38	1082	C	C	T	noncoding	n/d	n/a	n/a	n/a
45	7511726	4550605T1	SNP00092454	38	1084	C	C	T	noncoding	n/d	n/a	n/a	n/a
45	7511726	7627524J1	SNP00003186	204	283	T	T	C	V81	n/a	n/a	n/a	n/a
45	7511726	7627524J1	SNP00132159	348	427	T	T	C	L129	n/a	n/a	n/a	n/a
45	7511726	7627524J1	SNP00132160	453	532	C	C	G	A164	n/a	n/a	n/a	n/a
45	7511726	7707062J1	SNP00003185	13	74	T	C	T	A11	n/a	n/a	n/a	n/a
45	7511726	7707062J1	SNP00003186	203	266	C	T	C	I75	n/a	n/a	n/a	n/a
45	7511726	7707062J1	SNP00132158	6	67	T	T	C	F9	n/a	n/a	n/a	n/a
45	7511726	7707062J1	SNP00132159	349	410	T	T	C	L123	n/a	n/a	n/a	n/a
47	7511078	1461423H1	SNP00035960	42	790	C	C	A	L260	n/d	n/d	n/d	n/d
47	7511078	2564207H1	SNP00035961	82	1029	A	G	A	noncoding	n/a	n/a	n/a	n/a
47	7511078	2960723F6	SNP00035958	186	502	C	C	T	L164	n/a	n/a	n/a	n/a
47	7511078	2960723T6	SNP00035960	362	788	C	C	A	A259	n/d	n/d	n/d	n/d
47	7511078	2960723T6	SNP00035961	123	1027	G	G	A	noncoding	n/a	n/a	n/a	n/a
47	7511078	4110735H1	SNP00035959	130	657	A	A	G	A215	n/a	n/a	n/a	n/a
47	7511078	5652256H1	SNP00035958	240	506	C	C	T	P165	n/a	n/a	n/a	n/a
47	7511078	5652256H1	SNP00035960	528	794	C	C	A	P261	n/d	n/d	n/d	n/d
48	7511505	031545H1	SNP00011025	69	1403	T	T	C	G392	n/a	n/a	n/a	n/a
48	7511505	1269946T6	SNP00011026	423	2040	C	C	T	noncoding	0.57	0.51	0.44	0.7
48	7511505	1288103H1	SNP00054030	244	866	C	C	T	I213	n/a	n/a	n/a	n/a
48	7511505	1288103H1	SNP00137093	219	841	T	T	C	L205	n/a	n/a	n/a	n/a
48	7511505	1377666F6	SNP00041731	23	1678	G	G	A	noncoding	n/a	n/a	n/a	n/a
48	7511505	1377666T6	SNP00041731	252	1685	G	G	A	noncoding	n/a	n/a	n/a	n/a
48	7511505	1477055H1	SNP00054031	217	1289	C	C	T	D354	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
48	7511505	1649779T6	SNP00011026	294	2162	C	C	T	noncoding	n/a	n/a	n/a	n/a
48	7511505	1737071H1	SNP00011026	214	2039	C	C	T	noncoding	0.57	0.51	0.44	0.7
48	7511505	2484384H1	SNP00041727	205	205	T	T	C	noncoding	n/a	n/a	n/a	n/a
48	7511505	2512592F6	SNP00041728	49	319	G	A	G	R31	n/a	n/a	n/a	n/a
48	7511505	2512592F6	SNP00041729	91	362	T	C	T	T45	n/a	n/a	n/a	n/a
48	7511505	2512592F6	SNP00041730	298	569	C	C	T	N114	0.98	0.87	0.9	0.86
48	7511505	3518380T6	SNP00011026	298	2161	C	C	T	noncoding	n/a	n/a	n/a	n/a
48	7511505	7255731H2	SNP00041728	13	320	A	A	G	R31	n/a	n/a	n/a	n/a
48	7511505	7650249J2	SNP00054030	248	867	C	C	T	L214	n/a	n/a	n/a	n/a
48	7511505	7650249J2	SNP00137093	273	842	T	T	C	L205	n/a	n/a	n/a	n/a
48	7511505	7763246J1	SNP00041728	98	294	A	A	G	T23	n/a	n/a	n/a	n/a
48	7511505	7763246J1	SNP00041729	140	337	C	C	T	T37	n/a	n/a	n/a	n/a
48	7511505	7763246J1	SNP00041730	347	544	C	C	T	A106	0.98	0.87	0.9	0.86
49	7511552	1224235H1	SNP00011392	226	424	T	T	C	S75	n/a	n/a	n/a	n/a
49	7511552	1284532H1	SNP00042220	176	659	T	T	C	Y154	n/a	n/a	n/a	n/a
49	7511552	7455113H2	SNP00011392	185	391	T	T	C	V64	n/a	n/a	n/a	n/a
49	7511552	7455113H2	SNP00042220	421	627	T	T	C	L143	n/a	n/a	n/a	n/a
50	7511722	1210857R1	SNP00008678	55	477	C	C	T	D158	n/a	n/a	n/a	n/a
50	7511722	1709352F6	SNP00008678	316	466	C	C	T	R155	n/a	n/a	n/a	n/a
50	7511722	1709352H1	SNP00132026	140	290	A	A	G	Y96	n/a	n/a	n/a	n/a
50	7511722	1709764T6	SNP00105006	20	791	T	T	G	noncoding	n/d	0.99	n/d	n/d
50	7511722	1710463H1	SNP00035683	113	262	C	C	T	R87	n/d	n/d	n/d	n/d
50	7511722	1841101H1	SNP00105006	260	810	T	T	G	noncoding	n/d	0.99	n/d	n/d
50	7511722	1841101R6	SNP00105006	257	807	T	T	G	noncoding	n/d	0.99	n/d	n/d
50	7511722	1841101T6	SNP00105006	10	790	T	T	G	noncoding	n/d	0.99	n/d	n/d
50	7511722	6214728H1	SNP00058604	279	582	G	G	A	E193	n/a	n/a	n/a	n/a
50	7511722	6216796H1	SNP00096536	399	422	T	T	G	L140	n/d	n/d	n/d	n/d
50	7511722	6597824F8	SNP00120193	313	417	C	G	C	G138	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
50	7511722	6738488H1	SNP00140687	273	615	A	A	G	A204	n/a	n/a	n/a	n/a
51	7511489	000278H1	SNP00123117	232	1136	T	T	C	L333	n/d	n/a	n/a	n/a
51	7511489	020177H1	SNP00003188	221	932	G	C	G	L265	n/a	n/a	n/a	n/a
51	7511489	1274759T6	SNP00003189	63	1449	T	T	C	noncoding	n/a	n/a	n/a	n/a
51	7511489	1291827H1	SNP00123116	61	683	C	C	T	G182	n/a	n/a	n/a	n/a
51	7511489	1624459F6	SNP00065489	216	271	A	A	G	K45	n/a	n/a	n/a	n/a
51	7511489	1624459F6	SNP00072434	103	158	T	C	T	L7	n/a	n/a	n/a	n/a
51	7511489	2112792R6	SNP00065489	190	277	A	A	G	H47	n/a	n/a	n/a	n/a
51	7511489	2112792R6	SNP00072434	77	164	C	C	T	H9	n/a	n/a	n/a	n/a
51	7511489	2112792T6	SNP00003188	224	993	G	C	G	V286	n/a	n/a	n/a	n/a
51	7511489	3120473T6	SNP00003189	19	1448	T	T	C	noncoding	n/a	n/a	n/a	n/a
51	7511489	4425660H1	SNP00152940	26	759	G	A	G	G208	n/a	n/a	n/a	n/a
51	7511489	4805658H1	SNP00065489	196	273	A	A	G	N46	n/a	n/a	n/a	n/a
51	7511489	4805658H1	SNP00072434	83	160	T	C	T	V8	n/a	n/a	n/a	n/a
51	7511489	493625R6	SNP00065489	165	272	A	A	G	K45	n/a	n/a	n/a	n/a
51	7511489	493625R6	SNP00072434	52	159	T	C	T	C8	n/a	n/a	n/a	n/a
51	7511489	493625T6	SNP00003189	110	1447	T	T	C	noncoding	n/a	n/a	n/a	n/a
51	7511489	5301835F6	SNP00152939	103	262	C	C	A	T42	n/a	n/a	n/a	n/a
51	7511489	551634T6	SNP00105147	238	991	C	C	T	A285	0.4	n/a	0.95	0.67
51	7511489	5773745H1	SNP00123116	167	684	C	C	T	L183	n/a	n/a	n/a	n/a
51	7511489	6799578J1	SNP00065489	55	261	A	A	G	T42	n/a	n/a	n/a	n/a
51	7511489	7243569H1	SNP00105147	63	967	C	C	T	A277	0.4	n/a	0.95	0.67
54	7511612	125846H1	SNP00154888	152	1403	A	A	C	noncoding	n/a	n/a	n/a	n/a
54	7511612	125846H1	SNP00154889	158	1409	A	A	C	noncoding	n/a	n/a	n/a	n/a
54	7511612	1259618R6	SNP00154890	62	1604	C	T	C	noncoding	n/a	n/a	n/a	n/a
54	7511612	1259618R6	SNP00154891	36	1630	A	A	G	noncoding	n/a	n/a	n/a	n/a
54	7511612	1669737F6	SNP00002753	10	576	G	G	C	noncoding	0.85	n/a	n/a	n/a
54	7511612	1939120T6	SNP00154890	8	1601	C	T	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
54	7511612	2101062T6	SNP00154888	65	1542	A	A	C	noncoding	n/a	n/a	n/a	n/a
54	7511612	2101062T6	SNP00154889	59	1548	A	A	C	noncoding	n/a	n/a	n/a	n/a
54	7511612	2993360F6	SNP00047294	186	1219	T	T	C	noncoding	0.98	n/a	n/a	n/a
54	7511612	3523116H1	SNP00154887	28	441	T	T	C	T113	n/a	n/a	n/a	n/a
54	7511612	3803815H1	SNP00002753	67	575	G	G	C	noncoding	0.85	n/a	n/a	n/a
54	7511612	6604874H1	SNP00128388	56	130	T	C	T	Y10	n/a	n/a	n/a	n/a
54	7511612	6604874H1	SNP00128389	68	142	C	T	C	P14	n/a	n/a	n/a	n/a
55	7511624	1273292F6	SNP00013069	22	961	G	G	T	G292	n/a	n/a	n/a	n/a
55	7511624	1334850F6	SNP00013067	60	90	G	G	T	A2	0.81	0.97	n/d	0.93
55	7511624	1508329F6	SNP00055468	113	740	T	T	C	C218	0.97	n/a	n/a	n/a
55	7511624	2636474H1	SNP00055466	209	299	C	C	T	N71	n/a	n/a	n/a	n/a
55	7511624	2636474H1	SNP00055467	224	314	G	G	T	V76	n/a	n/a	n/a	n/a
56	7511626	1273292F6	SNP00013069	22	994	G	G	T	G303	n/a	n/a	n/a	n/a
56	7511626	1334850F6	SNP00013067	60	90	G	G	T	A2	0.81	0.97	n/d	0.93
56	7511626	1508329F6	SNP00055468	113	773	T	T	C	C229	0.97	n/a	n/a	n/a
56	7511626	2568901H1	SNP00101209	232	435	C	C	A	P117	n/d	n/a	n/a	n/a
56	7511626	2636474H1	SNP00055466	209	299	C	C	T	N71	n/a	n/a	n/a	n/a
56	7511626	2636474H1	SNP00055467	224	314	G	G	T	V76	n/a	n/a	n/a	n/a
56	7511626	3014210H1	SNP00013068	188	519	A	G	A	S145	n/a	n/a	n/a	n/a
57	7512885	2373340T6	SNP00123257	388	653	A	A	G	Q214	n/a	n/a	n/a	n/a
57	7512885	3407727H1	SNP00123257	11	629	A	A	G	A206	n/a	n/a	n/a	n/a
58	7511965	1631132F6	SNP00012037	313	795	A	G	A	T169	0.78	n/a	n/a	n/a
58	7511965	3522462R6	SNP00012036	93	363	T	C	T	L25	n/d	n/a	n/a	n/a
58	7511965	3712320F6	SNP00044775	409	1156	G	G	A	D290	0.69	0.86	0.91	0.78
59	7512403	1004612T6	SNP00116769	122	1325	T	T	C	stop374	n/a	n/a	n/a	n/a
59	7512403	1256663F6	SNP00116769	111	1344	T	T	C	noncoding	n/a	n/a	n/a	n/a
59	7512403	1321148F6	SNP00058613	29	2447	T	T	C	noncoding	n/d	n/a	n/a	n/a
59	7512403	1321148F6	SNP00058614	33	2451	A	A	G	noncoding	n/d	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
59	7512403	1321148H1	SNP00058614	77	2495	A	A	G	noncoding	n/d	n/d	n/d	n/d
59	7512403	1816079F6	SNP00025602	232	1138	C	C	A	S311	n/a	n/a	n/a	n/a
59	7512403	1969119T6	SNP00058614	52	2497	A	A	G	noncoding	n/d	n/d	n/d	n/d
59	7512403	2962164T7	SNP00116769	59	1328	T	T	C	Y375	n/a	n/a	n/a	n/a
59	7512403	961757R6	SNP00116769	147	1350	T	T	C	noncoding	n/a	n/a	n/a	n/a
59	7512403	961757T6	SNP00116769	122	1342	T	T	C	noncoding	n/a	n/a	n/a	n/a
60	7512564	007977H1	SNP00006647	139	1639	C	T	C	noncoding	n/a	n/a	n/a	n/a
60	7512564	1385137F6	SNP00135814	114	1500	A	A	G	noncoding	n/a	n/a	n/a	n/a
60	7512564	1385138T6	SNP00006647	13	1635	T	T	C	noncoding	n/a	n/a	n/a	n/a
60	7512564	1385138T6	SNP00135814	151	1496	A	A	G	noncoding	n/a	n/a	n/a	n/a
60	7512564	4111577T6	SNP00149378	132	1411	C	C	T	noncoding	n/a	n/a	n/a	n/a
60	7512564	4594352F6	SNP00006646	61	1406	A	A	G	noncoding	n/a	n/a	n/a	n/a
60	7512564	5764290H1	SNP00006647	283	1644	T	T	C	noncoding	n/a	n/a	n/a	n/a
60	7512564	5764290H1	SNP00135814	145	1505	A	A	G	noncoding	n/a	n/a	n/a	n/a
61	7512646	2044743H1	SNP00048181	156	395	C	T	C	D124	n/a	n/a	n/a	n/a
63	7512707	3595491H1	SNP00002073	223	597	T	C	T	H190	n/a	n/a	n/a	n/a
63	7512707	5845804H1	SNP00149574	38	856	C	C	T	H277	n/a	n/a	n/a	n/a
64	7512710	1832824T6	SNP00002074	216	1114	G	G	A	D367	n/a	n/a	n/a	n/a
64	7512710	2079512T6	SNP00002074	248	1097	A	G	A	E361	n/a	n/a	n/a	n/a
64	7512710	2434429H1	SNP00002074	132	1096	G	G	A	V361	n/a	n/a	n/a	n/a
64	7512710	2654610H1	SNP00149574	90	709	C	C	T	H232	n/a	n/a	n/a	n/a
64	7512710	3152185T6	SNP00002074	156	1172	A	G	A	K386	n/a	n/a	n/a	n/a
64	7512710	3187495T6	SNP00002074	194	1129	A	G	A	K372	n/a	n/a	n/a	n/a
64	7512710	3595491H1	SNP00002073	223	450	T	C	T	H145	n/a	n/a	n/a	n/a
65	7512884	088573H1	SNP00009024	86	211	C	C	T	noncoding	n/a	n/a	n/a	n/a
65	7512884	1740626R6	SNP00036860	219	1595	C	C	T	noncoding	n/a	n/a	n/a	n/a
65	7512884	2050911H1	SNP00107326	53	1153	A	A	G	noncoding	n/a	n/a	n/a	n/a
65	7512884	3236119F6	SNP00009024	94	212	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
65	7512884	3247348H1	SNP00036859	246	855	T	T	C	noncoding	n/d	n/d	n/d	n/d
65	7512884	5512625F6	SNP00048574	485	808	T	C	T	noncoding	n/d	n/a	n/a	n/a
66	7512931	2257046H1	SNP00059656	13	735	T	T	C	I178	n/a	n/a	n/a	n/a
66	7512931	2615823H1	SNP00028270	6	52	G	G	T	noncoding	0.98	n/a	n/a	n/a
66	7512931	2615823H1	SNP00028271	29	75	C	C	G	noncoding	n/a	n/a	n/a	n/a
66	7512931	5970391H1	SNP00059657	17	835	G	T	G	R211	n/a	n/a	n/a	n/a
66	7512931	5970391H1	SNP00059658	455	1273	C	A	C	T357	n/a	n/a	n/a	n/a
67	7512933	2615823H1	SNP00028270	6	52	G	G	T	noncoding	0.98	n/a	n/a	n/a
67	7512933	2615823H1	SNP00028271	29	75	C	C	G	noncoding	n/a	n/a	n/a	n/a
68	7512942	2615823H1	SNP00028270	6	36	G	G	T	noncoding	0.98	n/a	n/a	n/a
68	7512942	2615823H1	SNP00028271	29	59	C	C	G	noncoding	n/a	n/a	n/a	n/a
70	7512779	181246H1	SNP00039384	160	372	C	C	G	A101	n/d	n/a	n/a	n/a
70	7512779	4275007H1	SNP00003729	257	1946	C	C	G	noncoding	0.97	n/d	n/d	n/d
70	7512779	6705322F6	SNP00119494	219	197	G	G	C	V43	n/a	n/a	n/a	n/a
70	7512779	6705322H1	SNP00119494	299	112	G	G	C	P14	n/a	n/a	n/a	n/a
70	7512779	7359891H1	SNP00119495	251	738	A	A	G	D223	n/a	n/a	n/a	n/a
71	7512877	1308817H1	SNP00058761	170	1249	T	T	G	noncoding	n/d	n/a	n/a	n/a
71	7512877	1441063T6	SNP00033300	111	1745	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1550727T6	SNP00033300	112	1739	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1550727T6	SNP00036178	373	1478	C	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1627381F6	SNP00036178	204	1449	C	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1627381F6	SNP00105496	158	1403	C	C	T	noncoding	n/d	n/d	n/d	n/d
71	7512877	1627381T6	SNP00033300	107	1760	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1627381T6	SNP00036178	368	1499	C	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1665743H1	SNP00075867	8	538	C	C	T	P90	n/a	n/a	n/a	n/a
71	7512877	1726046F6	SNP00036178	203	1448	T	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1726046F6	SNP00105496	157	1402	C	C	T	noncoding	n/d	n/d	n/d	n/d
71	7512877	1726046T6	SNP00033300	121	1729	G	G	T	noncoding	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
71	7512877	1729440T6	SNP00033300	115	1738	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1729440T6	SNP00036178	376	1477	C	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	1808254H1	SNP00033300	183	1710	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	781216T6	SNP00033300	135	1737	G	G	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	781216T6	SNP00036178	396	1476	C	C	T	noncoding	n/d	n/a	n/a	n/a
71	7512877	781216T6	SNP00105496	441	1431	C	C	T	noncoding	n/d	n/d	n/d	n/d
74	7512794	7585776H2	SNP00069120	359	354	T	T	G	W17	n/d	n/d	n/d	n/d
75	7512886	1752587T6	SNP00014852	30	1016	T	T	C	noncoding	n/d	n/d	n/d	n/d
75	7512886	2555028F6	SNP00014851	257	781	G	A	G	noncoding	n/a	n/a	n/a	n/a
75	7512886	391692R6	SNP00014852	138	1015	T	T	C	noncoding	n/d	n/d	n/d	n/d
76	7512929	1363313F6	SNP00151594	222	771	C	C	T	noncoding	n/a	n/a	n/a	n/a
76	7512929	1363313T6	SNP00028980	378	1420	T	T	C	noncoding	n/a	n/a	n/a	n/a
76	7512929	1364310T6	SNP00028980	384	1414	T	T	C	noncoding	n/a	n/a	n/a	n/a
76	7512929	1984112T6	SNP00028980	316	1481	T	T	C	noncoding	n/a	n/a	n/a	n/a
76	7512929	2109361H1	SNP00069824	129	339	C	C	T	T97	n/a	n/a	n/a	n/a
76	7512929	2909762H1	SNP00028980	17	1403	T	T	C	noncoding	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14-16, SEQ ID NO:30, SEQ ID NO:32-33, and SEQ ID NO:36-37,

c) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:9-10, SEQ ID NO:12-13, SEQ ID NO:17-19, SEQ ID NO:21-22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:31, SEQ ID NO:34-35, and SEQ ID NO:38,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:6,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:20, SEQ ID NO:25, and SEQ ID NO:29,

f) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:23 and SEQ ID NO:27,

g) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, and

h) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-76,

b) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:39-40, SEQ ID NO:45-48, SEQ ID NO:50, SEQ ID NO:52-53, SEQ ID NO:57, SEQ ID NO:59-69, and SEQ ID NO:74-76,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:41, SEQ ID NO:44, and SEQ ID NO:55-56,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

98% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:42 and SEQ ID NO:70,

e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to the polynucleotide sequence of SEQ ID NO:43,

f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91% identical to the polynucleotide sequence of SEQ ID NO:49,

g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 97% identical to the polynucleotide sequence of SEQ ID NO:51,

h) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:54 and SEQ ID NO:58,

i) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:71-73,

j) a polynucleotide complementary to a polynucleotide of a),

k) a polynucleotide complementary to a polynucleotide of b),

l) a polynucleotide complementary to a polynucleotide of c),

m) a polynucleotide complementary to a polynucleotide of d),

n) a polynucleotide complementary to a polynucleotide of e),

o) a polynucleotide complementary to a polynucleotide of f),

p) a polynucleotide complementary to a polynucleotide of g),

q) a polynucleotide complementary to a polynucleotide of h),

r) a polynucleotide complementary to a polynucleotide of i), and

s) an RNA equivalent of a)-r).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions

whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

19. A method for treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) contacting a sample comprising a polypeptide of claim 1 with a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of

functional ENZM, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) contacting a sample comprising a polypeptide of claim 1 with a compound, and
b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional ENZM, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) contacting a sample comprising the target polynucleotide with a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of screening for potential toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample indicates potential toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of ENZM in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of ENZM in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of ENZM in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-38, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence

selected from the group consisting of SEQ ID NO:1-38, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38 in a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38 from a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide

comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-38.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

91. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:36.

92. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:37.

93. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:56.

112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:57.

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:66.

122. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:67.

123. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:68.

124. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:69.

125. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:70.

126. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:71.

127. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:72.

128. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:73.

129. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:74.

130. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:75.

131. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:76.

```
<110> INCYTE CORPORATION
      LEE, Soo Y.
      KABLE, Amy E.
      CHAWLA, Narinder K.
      MASON, Patricia M.
      HAFALIA, April J.A.
      MARQUIS, Joseph P.
      RICHARDSON, Thomas W.
      ELLIOTT, Vicki S.
      BECHA, Shanya D.
      BULLOCH, Sean A.
      HAWKINS, Phillip R.
      JIN, Pei
      GIETZEN, Kimberly J.
      WILSON, Amy D.
      CHANG, Hsin-Ru
      KHARE, Reena
      SWARNAKAR, Anita
      TRAN, Uyen K.
      EMERLING, Brooke M.

<120> ENZYMES

<130> PF-1440 PCT

<140> To Be Assigned
<141> Herewith

<150> US 60/387,119
<151> 2002-06-07

<150> US 60/390,662
<151> 2002-06-21

<160> 76
<170> PERL Program

<210> 11
<211> 256
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511289CD1

<400> 1
Met Ser Gln Gly Ser Pro Gly Asp Trp Ala Pro Leu Asp Pro Thr
 1               5                  10                  15
Pro Gly Pro Pro Ala Ser Pro Asn Pro Phe Val His Glu Leu His
                20                  25                  30
Leu Ser Arg Leu Gln Arg Val Lys Phe Cys Leu Leu Gly Ala Leu
                35                  40                  45
Leu Ala Pro Ile Arg Val Leu Leu Ala Phe Ile Val Leu Phe Leu
                50                  55                  60
Leu Trp Pro Phe Ala Trp Leu Gln Val Ala Gly Leu Ser Glu Glu
```

65 70 75
Gln Leu Gln Glu Pro Ile Thr Gly Trp Arg Lys Thr Val Cys His
80 85 90
Asn Gly Val Leu Gly Leu Ser Arg Leu Leu Phe Phe Leu Leu Gly
95 100 105
Phe Leu Arg Ile Arg Val Arg Gly Gln Arg Ala Ser Arg Leu Gln
110 115 120
Ala Pro Val Leu Val Ala Ala Pro His Ser Thr Phe Phe Asp Pro
125 130 135
Ile Val Leu Leu Pro Cys Asp Leu Pro Lys Val Val Ser Arg Ala
140 145 150
Glu Asn Leu Ser Val Pro Val Ile Gly Ala Leu Leu Arg Phe Asn
155 160 165
Gln Ala Ile Leu Val Ser Arg His Asp Pro Ala Ser Arg Arg Arg
170 175 180
Val Val Glu Glu Val Arg Arg Arg Ala Thr Ser Gly Gly Lys Trp
185 190 195
Pro Gln Val Leu Phe Phe Pro Glu Gly Thr Cys Ser Asn Lys Lys
200 205 210
Ala Leu Leu Lys Phe Lys Pro Gly His His Gln Leu Gly Met Glu
215 220 225
Gly Ser Trp Ser Thr Gln Ser Pro Leu Ala His Ser Leu Ser Ala
230 235 240
Leu Gln His Cys Gly Cys Gly Val Pro Ser Cys Val Ser Pro Gln
245 250 255
Pro

<210> 2
<211> 138
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511056CD1

<400> 2
Met Val Phe Pro Ala Lys Arg Phe Cys Leu Val Pro Ser Met Glu
1 5 10 15
Gly Val Arg Trp Ala Phe Ser Cys Gly Thr Trp Leu Pro Ser Arg
20 25 30
Ala Glu Trp Leu Leu Ala Val Arg Ser Ile Gln Pro Glu Glu Lys
35 40 45
Glu Arg Ile Gly Gln Phe Val Phe Ala Arg Asp Ala Lys Ala Ala
50 55 60
Met Ala Gly Arg Leu Met Ile Arg Lys Leu Val Ala Glu Lys Leu
65 70 75
Asn Ile Pro Trp Asn His Ile Arg Leu Gln Arg Thr Ala Lys Gly
80 85 90
Lys Pro Val Leu Ala Lys Asp Ser Ser Asn Pro Tyr Pro Asn Phe
95 100 105
Asn Phe Asn Ile Ser His Gln Gly Asp Tyr Ala Val Leu Ala Ala
110 115 120
Glu Pro Glu Leu Gln Val Gly Ile Asp Ile Met Lys Thr Ser Phe
125 130 135

Pro Gly Thr

<210> 3
<211> 674
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511567CD1

<400> 3
```
Met Glu Thr Pro Ala Ala Ala Ala Pro Ala Gly Ser Leu Phe Pro
1               5                   10                  15
Ser Phe Leu Leu Leu Ala Cys Gly Thr Leu Val Ala Ala Leu Leu
                20                  25                  30
Gly Ala Ala His Arg Leu Gly Leu Phe Tyr Gln Leu Leu His Lys
                35                  40                  45
Val Asp Lys Ala Ser Val Arg His Gly Gly Glu Asn Val Ala Ala
                50                  55                  60
Val Leu Arg Ala His Gly Val Arg Phe Ile Phe Thr Leu Val Gly
                65                  70                  75
Gly His Ile Ser Pro Leu Leu Val Ala Cys Glu Lys Leu Gly Ile
                80                  85                  90
Arg Val Val Asp Thr Arg His Glu Val Thr Ala Val Phe Ala Ala
                95                  100                 105
Asp Ala Met Ala Arg Leu Ser Gly Thr Val Gly Val Ala Ala Val
                110                 115                 120
Thr Ala Gly Pro Gly Leu Thr Asn Thr Val Thr Ala Val Lys Asn
                125                 130                 135
Ala Gln Met Ala Gln Ser Pro Ile Leu Leu Leu Gly Gly Ala Ala
                140                 145                 150
Ser Thr Leu Leu Gln Asn Arg Gly Ala Leu Gln Ala Val Asp Gln
                155                 160                 165
Leu Ser Leu Phe Arg Pro Leu Cys Lys Phe Cys Val Ser Val Arg
                170                 175                 180
Arg Val Arg Asp Ile Val Pro Thr Leu Arg Ala Ala Met Ala Ala
                185                 190                 195
Ala Gln Ser Gly Thr Pro Gly Pro Val Phe Val Glu Leu Pro Val
                200                 205                 210
Asp Val Leu Tyr Pro Tyr Phe Met Val Gln Lys Glu Met Val Pro
                215                 220                 225
Ala Lys Pro Pro Lys Gly Leu Val Gly Arg Val Val Ser Trp Tyr
                230                 235                 240
Leu Glu Asn Tyr Leu Ala Asn Leu Phe Ala Gly Ala Trp Glu Pro
                245                 250                 255
Gln Pro Glu Gly Pro Leu Pro Leu Asp Ile Pro Gln Ala Ser Pro
                260                 265                 270
Gln Gln Val Gln Arg Cys Val Glu Ile Leu Ser Arg Ala Lys Arg
                275                 280                 285
Pro Leu Met Val Leu Gly Ser Gln Ala Leu Leu Thr Pro Thr Ser
                290                 295                 300
Ala Asp Lys Leu Arg Ala Ala Val Glu Thr Leu Gly Val Pro Cys
                305                 310                 315
Phe Leu Gly Gly Met Ala Arg Gly Leu Leu Gly Arg Asn His Pro
```

```
                320                 325                 330
Leu His Ile Arg Glu Asn Arg Ser Ala Ala Leu Lys Lys Ala Asp
                335                 340                 345
Val Ile Val Leu Ala Gly Thr Val Cys Asp Phe Arg Leu Ser Tyr
                350                 355                 360
Gly Arg Val Leu Ser His Ser Ser Lys Ile Ile Ile Val Asn Arg
                365                 370                 375
Asn Arg Glu Glu Met Leu Leu Asn Ser Asp Ile Phe Trp Lys Pro
                380                 385                 390
Gln Glu Ala Val Gln Gly Asp Val Gly Ser Phe Val Leu Lys Leu
                395                 400                 405
Val Glu Gly Leu Gln Gly Gln Thr Trp Ala Pro Asp Trp Val Glu
                410                 415                 420
Glu Leu Arg Glu Ala Asp Arg Gln Lys Glu Gln Thr Phe Arg Glu
                425                 430                 435
Lys Ala Ala Met Pro Val Ala Gln His Leu Asn Pro Val Gln Val
                440                 445                 450
Leu Gln Leu Val Glu Glu Thr Leu Pro Asp Asn Ser Ile Leu Val
                455                 460                 465
Val Asp Gly Gly Asp Phe Val Gly Thr Ala Ala His Leu Val Gln
                470                 475                 480
Pro Arg Gly Pro Leu Arg Trp Leu Asp Pro Gly Ala Phe Gly Thr
                485                 490                 495
Leu Gly Val Gly Ala Gly Phe Ala Leu Gly Ala Lys Leu Cys Arg
                500                 505                 510
Pro Asp Ala Glu Val Trp Cys Leu Phe Gly Asp Gly Ala Phe Gly
                515                 520                 525
Tyr Ser Leu Ile Glu Phe Asp Thr Phe Val Arg His Lys Ile Ile
                530                 535                 540
Thr Arg Gln Pro Trp Val Trp Gly Pro Gly Ala Cys Cys Ser His
                545                 550                 555
Gly Arg Thr Arg Ile Arg Trp Ser Arg Cys Cys Thr Met Pro Ser
                560                 565                 570
Ser Ser Ala Glu Thr Ala Thr Arg Leu Trp Ser Thr Ser Ser Leu
                575                 580                 585
Gly Gly Arg Thr Ser Ala Met Ala Pro Leu Leu Tyr Arg Ala Leu
                590                 595                 600
Trp Val Arg Thr Leu Gly Cys Leu Pro Ala Pro Gly Leu Cys Pro
                605                 610                 615
Ala Gly Leu Glu Ser His His Cys Leu Pro Trp Ala Tyr Pro Met
                620                 625                 630
Arg Pro Ser Asp Ser Thr Thr Ser Leu Arg Arg Gly Trp Arg Gly
                635                 640                 645
Gln Ile Ser Glu Arg Leu Ser Cys Glu Leu Leu Asp Pro Pro Leu
                650                 655                 660
His Gly Pro Asn Cys Ala Val Cys Pro Leu Ser Tyr Gly Asp
                665                 670
```

<210> 4
<211> 237
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511651CD1

<400> 4
Met Lys Val Leu Leu Leu Lys Asp Ala Lys Glu Asp Asp Cys Gly
1 5 10 15
Gln Asp Pro Tyr Ile Arg Leu Ser His Pro Glu Asp Tyr Gly Gly
20 25 30
Leu Ile Phe Thr Ser Pro Arg Ala Val Glu Ala Ala Glu Leu Cys
35 40 45
Leu Glu Gln Asn Asn Lys Thr Glu Val Trp Glu Arg Ser Leu Lys
50 55 60
Glu Lys Trp Asn Ala Lys Ser Val Tyr Val Val Gly Asn Ala Thr
65 70 75
Ala Ser Leu Val Ser Lys Ile Gly Leu Asp Thr Glu Gly Glu Thr
80 85 90
Cys Gly Asn Ala Glu Lys Leu Ala Glu Tyr Ile Cys Ser Arg Glu
95 100 105
Ser Ser Ala Leu Pro Leu Leu Phe Pro Cys Gly Asn Leu Lys Arg
110 115 120
Glu Ile Leu Pro Lys Ala Leu Lys Asp Lys Gly Ile Ala Met Glu
125 130 135
Ser Ile Thr Val Tyr Gln Thr Val Ala His Pro Gly Ile Gln Gly
140 145 150
Asn Leu Asn Ser Tyr Tyr Ser Gln Gln Gly Val Pro Ala Ser Ile
155 160 165
Thr Phe Phe Ser Pro Ser Gly Leu Thr Tyr Ser Leu Lys His Ile
170 175 180
Gln Glu Leu Ser Gly Asp Asn Ile Asp Gln Ile Lys Phe Ala Ala
185 190 195
Ile Gly Pro Thr Thr Ala Arg Ala Leu Ala Ala Gln Gly Leu Pro
200 205 210
Val Ser Cys Thr Ala Glu Ser Pro Thr Pro Gln Ala Leu Ala Thr
215 220 225
Gly Ile Arg Lys Ala Leu Gln Pro His Gly Cys Cys
230 235

<210> 5
<211> 178
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511881CD1

<400> 5
Met Lys Asp Glu Val Ala Leu Leu Ala Ala Val Thr Leu Leu Gly
1 5 10 15
Val Leu Leu Gln Ala Tyr Phe Ser Leu Gln Val Ile Ser Ala Arg
20 25 30
Arg Ala Phe Arg Val Ser Pro Pro Leu Thr Thr Gly Pro Pro Glu
35 40 45
Phe Glu Arg Val Tyr Arg Ala Gln Val Asn Cys Ser Glu Tyr Phe
50 55 60
Pro Leu Phe Leu Ala Thr Leu Trp Val Ala Gly Ile Phe Phe His
65 70 75
Glu Gly Arg Gly Val Gly Gln Gly Arg Thr Arg Trp Thr Pro Gly
80 85 90

```
Thr Arg Ala Gly Arg Ser Pro Gly Pro Cys Val Pro Leu Ala Gly
                95                  100                 105
Ala Ala Ala Leu Cys Gly Leu Val Tyr Leu Phe Ala Arg Leu Arg
                110                 115                 120
Tyr Phe Gln Gly Tyr Ala Arg Ser Ala Gln Leu Arg Leu Ala Pro
                125                 130                 135
Leu Tyr Ala Ser Ala Arg Ala Leu Trp Leu Leu Val Ala Leu Ala
                140                 145                 150
Ala Leu Gly Leu Leu Ala His Phe Leu Pro Ala Ala Leu Arg Ala
                155                 160                 165
Ala Leu Leu Gly Arg Leu Arg Thr Leu Leu Pro Trp Ala
                170                 175
```

<210> 6
<211> 245
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512181CD1

<400> 6

```
Met Gly Lys Leu Val Ala Leu Val Leu Leu Gly Val Gly Leu Ser
  1             5                   10                  15
Leu Val Gly Glu Met Phe Leu Ala Phe Arg Glu Arg Val Asn Ala
                20                  25                  30
Ser Arg Glu Val Glu Pro Val Glu Pro Glu Asn Cys His Leu Ile
                35                  40                  45
Glu Glu Leu Glu Ser Gly Ser Glu Asp Ile Asp Ile Leu Pro Ser
                50                  55                  60
Gly Leu Ala Phe Ile Ser Ser Gly Leu Lys Tyr Pro Gly Met Pro
                65                  70                  75
Asn Phe Ala Pro Asp Glu Pro Gly Lys Ile Phe Leu Met Asp Leu
                80                  85                  90
Asn Glu Gln Asn Pro Arg Ala Gln Ala Leu Glu Ile Ser Gly Gly
                95                  100                 105
Phe Asp Lys Glu Leu Phe Asn Pro His Gly Ile Ser Ile Phe Ile
                110                 115                 120
Asp Lys Asp Asn Thr Val Tyr Leu Tyr Val Val Asn His Pro His
                125                 130                 135
Met Lys Ser Thr Val Glu Ile Phe Lys Phe Glu Glu Gln Gln Arg
                140                 145                 150
Ser Leu Val Tyr Leu Lys Thr Ile Lys His Glu Leu Leu Lys Ser
                155                 160                 165
Val Asn Asp Ile Val Val Leu Gly Pro Glu Gln Phe Tyr Ala Thr
                170                 175                 180
Arg Asp His Tyr Phe Thr Asn Ser Leu Leu Ser Phe Phe Glu Met
                185                 190                 195
Ile Leu Asp Leu Arg Trp Thr Tyr Val Leu Phe Tyr Ser Pro Arg
                200                 205                 210
Glu Val Lys Val Val Ala Lys Gly Phe Cys Ser Ala Asn Gly Ile
                215                 220                 225
Thr Val Ser Ala Asp Gln Lys Tyr Phe Ala Ser Arg Met Phe Cys
                230                 235                 240
Leu Arg Ser Pro Gly
```

245

<210> 7
<211> 250
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511726CD1

<400> 7
Met Ala Ala Leu Arg Val Leu Leu Ser Cys Ala Arg Gly Pro Leu
1 5 10 15
Arg Pro Pro Val Arg Cys Pro Ala Trp Arg Pro Phe Ala Ser Gly
20 25 30
Ala Asn Phe Glu Tyr Ile Ile Ala Glu Lys Arg Gly Lys Asn Asn
35 40 45
Thr Val Gly Leu Ile Gln Leu Asn Arg Pro Lys Ala Leu Asn Ala
50 55 60
Leu Cys Asp Gly Leu Ile Asp Glu Leu Asn Gln Ala Leu Lys Ile
65 70 75
Phe Glu Glu Asp Pro Ala Val Gly Ala Ile Val Leu Thr Gly Gly
80 85 90
Asp Lys Ala Phe Ala Ala Gly Ala Asp Ile Lys Glu Met Gln Asn
95 100 105
Leu Ser Phe Gln Asp Cys Tyr Ser Ser Lys Phe Leu Lys His Trp
110 115 120
Asp His Leu Thr Gln Val Lys Lys Pro Val Ile Ala Ala Val Asn
125 130 135
Gly Tyr Ala Phe Gly Gly Gly Cys Glu Leu Ala Met Met Cys Asp
140 145 150
Ile Ile Tyr Ala Gly Glu Lys Ala Gln Phe Ala Gln Pro Glu Ile
155 160 165
Leu Ile Gly Thr Ile Pro Gly Ala Gly Gly Thr Gln Arg Leu Thr
170 175 180
Arg Ala Val Gly Lys Ser Leu Ala Met Glu Met Val Leu Thr Gly
185 190 195
Asp Arg Ile Ser Ala Gln Asp Ala Lys Gln Ala Ala Phe Glu Met
200 205 210
Thr Leu Thr Glu Gly Ser Lys Leu Glu Lys Lys Leu Phe Tyr Ser
215 220 225
Thr Phe Ala Thr Asp Asp Arg Lys Glu Gly Met Thr Ala Phe Val
230 235 240
Glu Lys Arg Lys Ala Asn Phe Lys Asp Gln
245 250

<210> 8
<211> 145
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511057CD1

<400> 8
Met Ala Lys Ser Leu Leu Lys Thr Ala Ser Leu Ser Gly Arg Thr
1 5 10 15
Lys Leu Leu His Gln Thr Gly Leu Ser Leu Tyr Ser Thr Ser His
20 25 30
Gly Phe Tyr Glu Glu Glu Val Lys Lys Thr Leu Gln Gln Phe Pro
35 40 45
Gly Gly Ser Ile Asp Leu Gln Lys Glu Asp Asn Gly Ile Gly Ile
50 55 60
Leu Thr Leu Asn Asn Pro Ser Arg Met Asn Ala Phe Ser Gly Val
65 70 75
Met Met Leu Gln Leu Leu Glu Lys Val Ile Glu Leu Glu Asn Trp
80 85 90
Thr Glu Gly Lys Gly Leu Ile Val Arg Gly Ala Lys Asn Thr Phe
95 100 105
Ser Ser Gly Ser Asp Leu Asn Ala Val Lys Ser Leu Gly Thr Pro
110 115 120
Glu Thr Ser Phe Asn Lys Cys Cys Ala Gly Ser Arg Leu Gly Ile
125 130 135
Gly Trp Arg Ser Arg Ile Tyr Tyr Ser Met
140 145

<210> 9
<211> 333
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511078CD1

<400> 9
Met Gly Ser Gln Val Ser Val Glu Ser Gly Ala Leu His Val Val
1 5 10 15
Ile Val Gly Gly Gly Phe Gly Gly Ile Ala Ala Ala Ser Gln Leu
20 25 30
Gln Ala Leu Asn Val Pro Phe Met Leu Val Asp Met Lys Asp Ser
35 40 45
Phe His His Asn Val Ala Ala Leu Arg Ala Ser Val Glu Thr Gly
50 55 60
Phe Ala Lys Lys Thr Phe Ile Ser Tyr Ser Val Thr Phe Lys Asp
65 70 75
Asn Phe Arg Gln Gly Leu Val Val Gly Ile Asp Leu Lys Asn Gln
80 85 90
Met Val Leu Leu Gln Gly Gly Glu Val Gln Arg Ser Arg Phe Ile
95 100 105
Val Val Val Gly Gly Gly Ser Ala Gly Val Glu Met Ala Ala Glu
110 115 120
Ile Lys Thr Glu Tyr Pro Glu Lys Glu Val Thr Leu Ile His Ser
125 130 135
Gln Val Ala Leu Ala Asp Lys Glu Leu Leu Pro Ser Val Arg Gln
140 145 150
Glu Val Lys Glu Ile Leu Leu Arg Lys Gly Val Gln Leu Leu Leu
155 160 165
Ser Glu Arg Val Ser Asn Leu Glu Glu Leu Pro Leu Asn Glu Tyr
170 175 180

Arg Glu Tyr Ile Lys Val Gln Thr Asp Lys Gly Thr Glu Val Ala
185 190 195
Thr Asn Leu Val Ile Leu Cys Thr Gly Ile Lys Ile Asn Ser Ser
200 205 210
Ala Tyr Arg Lys Ala Phe Glu Ser Arg Leu Ala Ser Ser Gly Ala
215 220 225
Leu Arg Val Asn Glu His Leu Gln Val Glu Gly His Ser Asn Val
230 235 240
Tyr Ala Ile Gly Asp Cys Ala Asp Val Arg Thr Pro Lys Met Ala
245 250 255
Tyr Leu Ala Gly Leu His Ala Asn Ile Ala Val Ala Asn Ile Val
260 265 270
Asn Ser Val Lys Gln Arg Pro Leu Gln Ala Tyr Lys Pro Gly Ala
275 280 285
Leu Thr Phe Leu Leu Ser Met Gly Arg Asn Asp Gly Val Gly Gln
290 295 300
Ile Ser Gly Phe Tyr Val Gly Arg Leu Met Val Arg Leu Thr Lys
305 310 315
Ser Arg Asp Leu Phe Val Ser Thr Ser Trp Lys Thr Met Arg Gln
320 325 330
Ser Pro Pro

<210> 10
<211> 443
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511505CD1

<400> 10
Met Ala Ala Lys Ser Gln Pro Asn Ile Pro Lys Ala Lys Ser Leu
1 5 10 15
Asp Gly Val Thr Asn Asp Arg Thr Ala Ser Gln Gly Gln Trp Gly
20 25 30
Arg Ala Cys Cys Ala Leu Thr Gly Pro Val Val Asp Ile Val Thr
35 40 45
Gly His Ala Arg Leu Ser Asp Ile Trp Ala Lys Thr Pro Pro Ile
50 55 60
Thr Arg Lys Ala Ala Gln Leu Tyr Thr Leu Trp Val Thr Phe Gln
65 70 75
Val Leu Leu Tyr Thr Ser Leu Pro Asp Phe Cys His Lys Phe Leu
80 85 90
Pro Gly Tyr Val Gly Gly Ile Gln Glu Gly Ala Val Thr Pro Ala
95 100 105
Gly Val Val Asn Lys Tyr Gln Ile Asn Gly Leu Gln Ala Trp Leu
110 115 120
Leu Thr His Leu Leu Trp Phe Ala Asn Ala His Leu Leu Ser Trp
125 130 135
Phe Ser Pro Thr Ile Ile Phe Asp Asn Trp Ile Pro Leu Leu Trp
140 145 150
Cys Ala Asn Ile Leu Gly Tyr Ala Val Ser Thr Phe Ala Met Val
155 160 165
Lys Gly Tyr Phe Phe Pro Thr Ser Ala Arg Asp Cys Lys Phe Thr

```
                170                 175                 180
Gly Asn Phe Phe Tyr Asn Tyr Met Met Gly Ile Glu Phe Asn Pro
                185                 190                 195
Arg Ile Gly Lys Trp Phe Asp Phe Lys Leu Phe Phe Asn Gly Arg
                200                 205                 210
Pro Gly Ile Val Ala Trp Thr Leu Ile Asn Leu Ser Phe Ala Ala
                215                 220                 225
Lys Gln Arg Glu Leu His Ser His Val Thr Asn Ala Met Val Leu
                230                 235                 240
Val Asn Val Leu Gln Ala Ile Tyr Val Ile Asp Phe Phe Trp Asn
                245                 250                 255
Glu Thr Trp Tyr Leu Lys Thr Ile Asp Ile Cys His Asp His Phe
                260                 265                 270
Gly Trp Tyr Leu Gly Trp Gly Asp Cys Val Trp Leu Pro Tyr Leu
                275                 280                 285
Tyr Thr Leu Gln Gly Leu Tyr Leu Val Tyr His Pro Val Gln Leu
                290                 295                 300
Ser Thr Pro His Ala Val Gly Val Leu Leu Leu Gly Leu Val Gly
                305                 310                 315
Tyr Tyr Ile Phe Arg Val Ala Asn His Gln Lys Asp Leu Phe Arg
                320                 325                 330
Arg Thr Asp Gly Arg Cys Leu Ile Trp Gly Arg Lys Pro Lys Val
                335                 340                 345
Ile Glu Cys Ser Tyr Thr Ser Ala Asp Gly Gln Arg His His Ser
                350                 355                 360
Lys Leu Leu Val Ser Gly Phe Trp Gly Val Ala Arg His Phe Asn
                365                 370                 375
Tyr Val Gly Asp Leu Met Gly Ser Leu Ala Tyr Cys Leu Ala Cys
                380                 385                 390
Gly Gly Gly His Leu Leu Pro Tyr Phe Tyr Ile Ile Tyr Met Ala
                395                 400                 405
Ile Leu Leu Thr His Arg Cys Leu Arg Asp Glu His Arg Cys Ala
                410                 415                 420
Ser Lys Tyr Gly Arg Asp Trp Glu Arg Tyr Thr Ala Ala Val Pro
                425                 430                 435
Tyr Arg Leu Leu Pro Gly Ile Phe
                440

<210> 11
<211> 247
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511552CD1

<400> 11
Met Ser Ser Ser Gly Thr Pro Asp Leu Pro Val Leu Leu Thr Asp
  1               5                  10                  15
Leu Lys Ile Gln Tyr Thr Lys Ile Phe Ile Asn Asn Glu Trp His
                 20                  25                  30
Asp Ser Val Ser Gly Lys Lys Phe Pro Val Phe Asn Pro Ala Thr
                 35                  40                  45
Glu Glu Glu Leu Cys Gln Val Glu Glu Gly Asp Lys Glu Asp Val
                 50                  55                  60
```

Asp Lys Ala Val Lys Ala Ala Arg Gln Ala Phe Gln Ile Gly Ser
65 70 75
Pro Trp Arg Thr Met Asp Ala Ser Glu Arg Gly Arg Leu Leu Tyr
80 85 90
Lys Leu Ala Asp Leu Ile Glu Arg Asp Arg Leu Leu Leu Ala Thr
95 100 105
Met Glu Ser Met Asn Gly Gly Lys Leu Tyr Ser Asn Ala Tyr Leu
110 115 120
Asn Asp Leu Ala Gly Cys Ile Lys Thr Leu Arg Tyr Cys Ala Gly
125 130 135
Trp Ala Asp Lys Ile Gln Gly Arg Thr Ile Pro Ile Asp Gly Asn
140 145 150
Phe Phe Thr Tyr Thr Arg His Glu Pro Ile Gly Val Cys Gly Gln
155 160 165
Ile Ile Pro Trp Asn Phe Pro Leu Val Met Leu Ile Trp Lys Ile
170 175 180
Gly Pro Ala Leu Ser Cys Gly Asn Thr Val Val Val Lys Pro Ala
185 190 195
Glu Gln Thr Pro Leu Thr Ala Leu His Val Ala Ser Leu Ile Lys
200 205 210
Glu Ala Arg Gly Gly Ile Lys Ala Thr Leu Ser Ser Pro Gln Cys
215 220 225
Ser Leu Met Leu Gln Met Arg Cys Ala Leu Pro Lys Arg Arg Phe
230 235 240
Leu Asp Gln Cys Ser Lys Ser
245

<210> 12
<211> 237
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511722CD1

<400> 12
Met Ala Pro Leu Arg Phe Ser Ala Asn Leu Ser Trp Leu Phe Pro
1 5 10 15
Asp Leu Ser Gly Leu Pro Ala Arg Val Arg Ala Ala Gly Ser Ser
20 25 30
Gly Phe Glu Ala Val Glu Val Ala Trp Pro Tyr Ala Glu Thr Pro
35 40 45
Glu Ala Leu Ala Arg Ala Ala Arg Glu Ala Gly Leu Arg Leu Val
50 55 60
Leu Ile Asn Thr Pro Pro Gly Asp Gln Glu Lys Gly Glu Met Gly
65 70 75
Leu Gly Ala Val Pro Gly Arg Gln Ala Ala Phe Arg Glu Gly Leu
80 85 90
Glu Gln Ala Val Arg Tyr Ala Lys Ala Leu Gly Cys Pro Arg Ile
95 100 105
His Leu Met Ala Gly Arg Val Pro Gln Gly Ala Asp Arg Ile Ala
110 115 120
Val Lys Ala Glu Met Glu Ala Val Phe Leu Glu Asn Leu Arg His
125 130 135
Ala Ala Gly Val Leu Ala Gln Glu Asp Leu Val Gly Leu Leu Glu

140 145 150
Pro Ile Asn Thr Arg Ile Thr Asp Pro Gln Tyr Phe Leu Asp Thr
155 160 165
Pro Gln Gln Ala Ala Ala Ile Leu Gln Lys Asp Ile Phe His Trp
170 175 180
Gln Ile Met Asp Gly Asn Leu Thr Gly Asn Ile Arg Glu Phe Leu
185 190 195
Pro Ile Val Gly His Val Gln Val Ala Gln Val Pro Gly Arg Gly
200 205 210
Glu Pro Ser Ser Pro Gly Glu Leu Asn Phe Pro Tyr Leu Phe Gln
215 220 225
Leu Leu Glu Asp Glu Gly Tyr Lys Gly Phe Val Gly
230 235

<210> 13
<211> 358
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511489CD1

<400> 13
Met Glu Phe Val Lys Cys Leu Gly His Pro Glu Glu Phe Tyr Asn
1 5 10 15
Leu Val Arg Phe Arg Ile Gly Gly Lys Arg Lys Val Met Pro Lys
20 25 30
Met Asp Gln Asp Ser Leu Ser Ser Ser Leu Lys Thr Cys Tyr Lys
35 40 45
Tyr Leu Asn Gln Thr Ser Arg Ser Phe Ala Ala Val Ile Gln Ala
50 55 60
Leu Asp Gly Glu Met Arg Asn Ala Val Cys Ile Phe Tyr Leu Val
65 70 75
Leu Arg Ala Leu Asp Thr Leu Glu Asp Asp Met Thr Ile Ser Val
80 85 90
Glu Lys Lys Val Pro Leu Leu His Asn Phe His Ser Phe Leu Tyr
95 100 105
Gln Pro Asp Trp Arg Phe Met Glu Ser Lys Glu Lys Asp Arg Gln
110 115 120
Val Leu Glu Asp Phe Pro Thr Ile Ser Leu Glu Phe Arg Asn Leu
125 130 135
Ala Glu Lys Tyr Gln Thr Val Ile Ala Asp Ile Cys Arg Arg Met
140 145 150
Gly Ile Gly Met Ala Glu Phe Leu Asp Lys His Val Thr Ser Glu
155 160 165
Gln Glu Trp Asp Lys Tyr Cys His Tyr Val Ala Gly Leu Val Gly
170 175 180
Ile Gly Leu Ser Arg Leu Phe Ser Ala Ser Glu Phe Glu Asp Pro
185 190 195
Leu Val Gly Glu Asp Thr Glu Arg Ala Asn Ser Met Gly Leu Phe
200 205 210
Leu Gln Lys Thr Asn Ile Ile Arg Asp Tyr Leu Glu Asp Gln Gln
215 220 225
Gly Gly Arg Glu Phe Trp Pro Gln Glu Val Met Ala Ile Ala Thr
230 235 240

```
Leu Ala Ala Cys Tyr Asn Asn Gln Gln Val Phe Lys Gly Ala Val
                245                 250                 255
Lys Ile Arg Lys Gly Gln Ala Val Thr Leu Met Met Asp Ala Thr
                260                 265                 270
Asn Met Pro Ala Val Lys Ala Ile Ile Tyr Gln Tyr Met Glu Glu
                275                 280                 285
Ile Tyr His Arg Ile Pro Asp Ser Asp Pro Ser Ser Ser Lys Thr
                290                 295                 300
Arg Gln Ile Ile Ser Thr Ile Arg Thr Gln Asn Leu Pro Asn Cys
                305                 310                 315
Gln Leu Ile Ser Arg Ser His Tyr Ser Pro Ile Tyr Leu Ser Phe
                320                 325                 330
Val Met Leu Leu Ala Ala Leu Ser Trp Gln Tyr Leu Thr Thr Leu
                335                 340                 345
Ser Gln Val Thr Glu Asp Tyr Val Gln Thr Gly Glu His
                350                 355

<210> 14
<211> 140
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511497CD1

<400> 14
Met Ala Ser Arg Gly Arg Arg Pro Glu His Gly Gly Pro Pro Glu
  1               5                  10                  15
Leu Phe Tyr Asp Glu Thr Glu Ala Arg Lys Tyr Val Arg Asn Ser
                 20                  25                  30
Arg Met Ile Asp Ile Gln Thr Arg Met Ala Gly Arg Ala Leu Glu
                 35                  40                  45
Leu Leu Tyr Leu Pro Glu Asn Lys Pro Cys Tyr Leu Leu Asp Ile
                 50                  55                  60
Gly Cys Gly Thr Gly Leu Ser Gly Ser Tyr Leu Ser Asp Glu Gly
                 65                  70                  75
His Tyr Trp Val Gly Leu Asp Ile Ser Pro Ala Met Leu Ala Phe
                 80                  85                  90
Leu Leu Cys Ser Gly Ser Val Met Leu Thr Arg Ser Leu Lys Thr
                 95                 100                 105
Leu Pro Ser Ala Cys Thr Ala Phe Leu Leu Leu Phe Phe Leu Phe
                110                 115                 120
Ser Ser Gly Asp Pro Glu Leu Ser Cys Ser Cys Thr Leu Arg Thr
                125                 130                 135
Gln Ser Ser Trp Ser
                140

<210> 15
<211> 248
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511498CD1
```

<400> 15
Met Ala Ser Arg Gly Arg Arg Pro Glu His Gly Gly Pro Pro Glu
1   5   10   15
Leu Phe Tyr Asp Glu Thr Glu Ala Arg Lys Tyr Val Arg Asn Ser
                20   25   30
Arg Met Ile Asp Ile Gln Thr Arg Met Ala Gly Arg Ala Leu Glu
                35   40   45
Leu Leu Tyr Leu Pro Glu Asn Lys Pro Cys Tyr Leu Leu Asp Ile
                50   55   60
Gly Cys Gly Thr Gly Leu Ser Gly Ser Tyr Leu Ser Asp Glu Gly
                65   70   75
His Tyr Trp Val Gly Leu Asp Ile Ser Pro Ala Met Leu Asp Glu
                80   85   90
Ala Val Asp Arg Glu Ile Glu Gly Asp Leu Leu Leu Gly Asp Met
                95   100   105
Gly Gln Gly Ile Pro Phe Lys Pro Gly Thr Phe Asp Gly Cys Ile
                110   115   120
Ser Ile Ser Ala Val Gln Trp Leu Cys Asn Ala Asn Lys Lys Ser
                125   130   135
Glu Asn Pro Ala Lys Arg Leu Tyr Cys Phe Phe Ala Ser Leu Phe
                140   145   150
Ser Val Leu Val Arg Gly Ser Arg Ala Val Leu Gln Leu Tyr Pro
                155   160   165
Glu Asn Ser Glu Gln Leu Glu Leu Ile Thr Thr Gln Ala Thr Lys
                170   175   180
Ala Gly Phe Ser Gly Gly Met Val Val Asp Tyr Pro Asn Ser Ala
                185   190   195
Lys Ala Lys Lys Phe Tyr Leu Cys Leu Phe Ser Gly Pro Ser Thr
                200   205   210
Phe Ile Pro Glu Val Pro Ile Lys Asp Val Glu Ala Gly Asn Gly
                215   220   225
Glu Glu Glu Ser Gly Met Gly Ala Gly Glu Glu Gly Ala Ala Gln
                230   235   240
Ala Pro Gly Gln Gly Ser Gln Thr
                245

<210> 16
<211> 153
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511612CD1

<400> 16
Met Gln Arg Leu Gln Val Val Leu Gly His Leu Arg Gly Pro Ala
1   5   10   15
Asp Ser Gly Trp Met Pro Gln Ala Ala Pro Cys Leu Ser Gly Ala
                20   25   30
Pro Gln Ala Ser Ala Ala Asp Val Val Val Val His Gly Arg Arg
                35   40   45
Thr Ala Ile Cys Arg Ala Gly Arg Gly Gly Phe Lys Asp Thr Thr
                50   55   60
Pro Asp Glu Leu Leu Ser Ala Val Met Thr Ala Val Leu Lys Asp
                65   70   75

```
Val Asn Leu Arg Pro Glu Gln Leu Gly Asp Ile Cys Val Gly Asn
                80                  85                  90
Val Leu Gln Pro Gly Ala Gly Ala Ile Met Ala Arg Ile Ala Gln
                95                  100                 105
Phe Leu Ser Asp Ile Pro Glu Thr Val Pro Leu Ser Thr Val Asn
                110                 115                 120
Arg Gln Cys Ser Ser Gly Leu Gln Ala Val Ala Ser Ile Ala Gly
                125                 130                 135
Trp Ser Pro Cys Pro Trp Leu Thr Glu Gly Thr Leu Glu Ile Leu
                140                 145                 150
Leu Arg Ala


<210> 17
<211> 376
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511624CD1

<400> 17
Met Ala Leu Leu His Ser Gly Arg Val Leu Pro Gly Ile Ala Ala
  1               5                   10                  15
Ala Phe His Pro Gly Leu Ala Ala Ala Ala Ser Ala Arg Ala Ser
                20                  25                  30
Ser Trp Trp Thr His Val Glu Met Gly Pro Pro Asp Pro Ile Leu
                35                  40                  45
Gly Val Thr Glu Ala Phe Lys Arg Asp Thr Asn Ser Lys Lys Met
                50                  55                  60
Asn Leu Gly Val Gly Ala Tyr Arg Asp Asp Asn Gly Lys Pro Tyr
                65                  70                  75
Val Leu Pro Ser Val Arg Lys Ala Glu Ala Gln Ile Ala Ala Lys
                80                  85                  90
Asn Leu Asp Lys Glu Tyr Leu Pro Ile Gly Gly Leu Ala Glu Phe
                95                  100                 105
Cys Lys Ala Ser Ala Glu Leu Ala Leu Gly Glu Asn Ser Glu Val
                110                 115                 120
Leu Lys Ser Gly Arg Phe Val Thr Val Gln Thr Ile Ser Gly Thr
                125                 130                 135
Gly Ala Leu Arg Ile Gly Ala Ser Phe Leu Lys Ile Pro Glu Gln
                140                 145                 150
Ser Val Leu Leu Leu His Ala Cys Ala His Asn Pro Thr Gly Val
                155                 160                 165
Asp Pro Arg Pro Glu Gln Trp Lys Glu Ile Ala Thr Val Val Lys
                170                 175                 180
Lys Arg Asn Leu Phe Ala Phe Phe Asp Met Ala Tyr Gln Gly Phe
                185                 190                 195
Ala Ser Gly Asp Gly Asp Lys Asp Ala Trp Ala Val Arg His Phe
                200                 205                 210
Ile Glu Gln Gly Ile Asn Val Cys Leu Cys Gln Ser Tyr Ala Lys
                215                 220                 225
Asn Met Gly Leu Tyr Gly Glu Arg Val Gly Ala Phe Thr Met Val
                230                 235                 240
Cys Lys Asp Ala Asp Glu Ala Lys Arg Val Glu Ser Gln Leu Lys
```

```
                245                 250                 255
Ile Leu Ile Arg Pro Met Tyr Ser Asn Pro Pro Leu Asn Gly Ala
                260                 265                 270
Arg Ile Ala Ala Ala Ile Leu Asn Thr Pro Asp Leu Arg Lys Gln
                275                 280                 285
Trp Leu Gln Glu Val Lys Gly Met Ala Asp Arg Ile Ile Gly Met
                290                 295                 300
Arg Thr Gln Leu Val Ser Asn Leu Lys Lys Glu Gly Ser Thr His
                305                 310                 315
Asn Trp Gln His Ile Thr Asp Gln Ile Gly Met Phe Cys Phe Thr
                320                 325                 330
Gly Leu Lys Pro Glu Gln Val Glu Arg Leu Ile Lys Glu Phe Ser
                335                 340                 345
Ile Tyr Met Thr Lys Asp Gly Arg Ile Ser Val Ala Gly Val Thr
                350                 355                 360
Ser Ser Asn Val Gly Tyr Leu Ala His Ala Ile His Gln Val Thr
                365                 370                 375
Lys


<210> 18
<211> 387
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511626CD1

<400> 18
Met Ala Leu Leu His Ser Gly Arg Val Leu Pro Gly Ile Ala Ala
  1             5                   10                  15
Ala Phe His Pro Gly Leu Ala Ala Ala Ala Ser Ala Arg Ala Ser
                20                  25                  30
Ser Trp Trp Thr His Val Glu Met Gly Pro Pro Asp Pro Ile Leu
                35                  40                  45
Gly Val Thr Glu Ala Phe Lys Arg Asp Thr Asn Ser Lys Lys Met
                50                  55                  60
Asn Leu Gly Val Gly Ala Tyr Arg Asp Asp Asn Gly Lys Pro Tyr
                65                  70                  75
Val Leu Pro Ser Val Arg Lys Phe Val Thr Val Gln Thr Ile Ser
                80                  85                  90
Gly Thr Gly Ala Leu Arg Ile Gly Ala Ser Phe Leu Gln Arg Phe
                95                  100                 105
Phe Lys Phe Ser Arg Asp Val Phe Leu Pro Lys Pro Thr Trp Gly
                110                 115                 120
Asn His Thr Pro Ile Phe Arg Asp Ala Gly Met Gln Leu Gln Gly
                125                 130                 135
Tyr Arg Tyr Tyr Asp Pro Lys Thr Cys Gly Phe Asp Phe Thr Gly
                140                 145                 150
Ala Val Glu Asp Ile Ser Lys Ile Pro Glu Gln Ser Val Leu Leu
                155                 160                 165
Leu His Ala Cys Ala His Asn Pro Thr Gly Val Asp Pro Arg Pro
                170                 175                 180
Glu Gln Trp Lys Glu Ile Ala Thr Val Val Lys Lys Arg Asn Leu
                185                 190                 195
```

Phe Ala Phe Phe Asp Met Ala Tyr Gln Gly Phe Ala Ser Gly Asp
200 205 210
Gly Asp Lys Asp Ala Trp Ala Val Arg His Phe Ile Glu Gln Gly
215 220 225
Ile Asn Val Cys Leu Cys Gln Ser Tyr Ala Lys Asn Met Gly Leu
230 235 240
Tyr Gly Glu Arg Val Gly Ala Phe Thr Met Val Cys Lys Asp Ala
245 250 255
Asp Glu Ala Lys Arg Val Glu Ser Gln Leu Lys Ile Leu Ile Arg
260 265 270
Pro Met Tyr Ser Asn Pro Pro Leu Asn Gly Ala Arg Ile Ala Ala
275 280 285
Ala Ile Leu Asn Thr Pro Asp Leu Arg Lys Gln Trp Leu Gln Glu
290 295 300
Val Lys Gly Met Ala Asp Arg Ile Ile Gly Met Arg Thr Gln Leu
305 310 315
Val Ser Asn Leu Lys Lys Glu Gly Ser Thr His Asn Trp Gln His
320 325 330
Ile Thr Asp Gln Ile Gly Met Phe Cys Phe Thr Gly Leu Lys Pro
335 340 345
Glu Gln Val Glu Arg Leu Ile Lys Glu Phe Ser Ile Tyr Met Thr
350 355 360
Lys Asp Gly Arg Ile Ser Val Ala Gly Val Thr Ser Ser Asn Val
365 370 375
Gly Tyr Leu Ala His Ala Ile His Gln Val Thr Lys
380 385

<210> 19
<211> 274
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512885CD1

<400> 19
Met Val Ser Pro Ala Thr Arg Lys Ser Leu Pro Lys Val Lys Ala
1 5 10 15
Met Asp Phe Ile Thr Ser Thr Ala Ile Leu Pro Leu Leu Phe Gly
20 25 30
Cys Leu Gly Val Phe Gly Leu Phe Arg Leu Leu Gln Trp Val Arg
35 40 45
Gly Lys Ala Tyr Leu Arg Asn Ala Val Val Val Ile Thr Gly Ala
50 55 60
Thr Ser Gly Leu Gly Lys Glu Cys Ala Lys Val Phe Tyr Ala Ala
65 70 75
Gly Ala Lys Leu Val Leu Cys Gly Arg Asn Gly Gly Ala Leu Glu
80 85 90
Glu Leu Ile Arg Glu Leu Thr Ala Ser His Ala Thr Lys Val Gln
95 100 105
Thr His Lys Pro Tyr Leu Val Thr Phe Asp Leu Thr Asp Ser Gly
110 115 120
Ala Ile Val Ala Ala Ala Ala Glu Ile Leu Gln Cys Phe Gly Tyr
125 130 135
Val Asp Ile Leu Val Asn Asn Ala Gly Ile Ser Tyr Arg Gly Thr

```
                    140                      145                     150
Ile Met Asp Thr Thr Val Asp Val Asp Lys Arg Val Met Glu Thr
                    155                      160                     165
Asn Tyr Phe Gly Pro Val Ala Leu Thr Lys Ala Leu Leu Pro Ser
                    170                      175                     180
Met Ile Lys Arg Arg Gln Gly His Ile Val Ala Ile Ser Ser Ile
                    185                      190                     195
Gln Gly Lys Met Ser Ile Pro Phe Arg Ser Ala Phe Met Asp Thr
                    200                      205                     210
Thr Thr Ala Gln Gly Arg Ser Pro Val Glu Val Ala Gln Asp Val
                    215                      220                     225
Leu Ala Ala Val Gly Lys Lys Lys Lys Asp Val Ile Leu Ala Asp
                    230                      235                     240
Leu Leu Pro Ser Leu Ala Val Tyr Leu Arg Thr Leu Ala Pro Gly
                    245                      250                     255
Leu Phe Phe Ser Leu Met Ala Ser Arg Ala Arg Lys Glu Arg Lys
                    260                      265                     270
Ser Lys Asn Ser


&lt;210&gt; 20
&lt;211&gt; 293
&lt;212&gt; PRT
&lt;213&gt; Homo sapiens

&lt;220&gt;
&lt;221&gt; misc_feature
&lt;223&gt; Incyte ID No: 7511965CD1

&lt;400&gt; 20
Met Ala Thr Phe Val Glu Leu Ser Thr Lys Ala Lys Met Pro Ile
 1               5                   10                  15
Val Gly Leu Gly Thr Trp Arg Ser Leu Leu Gly Lys Val Lys Glu
                20                   25                  30
Ala Val Lys Val Ala Ile Asp Ala Gly Tyr Arg His Ile Asp Cys
                35                   40                  45
Ala Tyr Val Tyr Gln Asn Glu His Glu Val Gly Glu Ala Ile Gln
                50                   55                  60
Glu Lys Ile Gln Glu Lys Ala Val Met Arg Glu Asp Leu Phe Ile
                65                   70                  75
Val Ser Lys Val Trp Pro Thr Phe Phe Glu Arg Pro Leu Val Arg
                80                   85                  90
Lys Ala Phe Glu Lys Thr Leu Lys Asp Leu Lys Leu Ser Tyr Leu
                95                  100                 105
Asp Val Tyr Leu Ile His Trp Pro Gln Gly Phe Lys Thr Gly Asp
               110                  115                 120
Asp Phe Phe Pro Lys Asp Asp Lys Gly Asn Met Ile Ser Gly Lys
               125                  130                 135
Gly Thr Phe Leu Asp Ala Trp Glu Ile Glu Arg Leu Leu Asn Lys
               140                  145                 150
Pro Gly Leu Lys Tyr Lys Pro Val Thr Asn Gln Val Glu Cys His
               155                  160                 165
Pro Tyr Leu Thr Gln Glu Lys Leu Ile Gln Tyr Cys His Ser Lys
               170                  175                 180
Gly Ile Thr Val Thr Ala Tyr Ser Pro Leu Gly Ser Pro Asp Arg
               185                  190                 195
```

Pro Trp Ala Lys Pro Glu Asp Pro Ser Leu Leu Glu Asp Pro Lys
200 205 210
Ile Lys Glu Ile Ala Ala Lys His Lys Lys Thr Thr Ala Gln Val
215 220 225
Leu Ile Arg Phe His Ile Gln Arg Asn Val Thr Val Ile Pro Lys
230 235 240
Ser Met Thr Pro Ala His Ile Val Glu Asn Ile Gln Val Phe Asp
245 250 255
Phe Lys Leu Ser Asp Glu Glu Met Ala Thr Ile Leu Ser Phe Asn
260 265 270
Arg Asn Trp Arg Ala Phe Asp Phe Lys Glu Phe Ser His Leu Glu
275 280 285
Asp Phe Pro Phe Asp Ala Glu Tyr
290

<210> 21
<211> 376
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512403CD1

<400> 21
Met Glu Leu Glu Val Arg Arg Val Arg Gln Ala Phe Leu Ser Gly
1 5 10 15
Arg Ser Arg Pro Leu Arg Phe Arg Leu Gln Gln Leu Glu Ala Leu
20 25 30
Arg Arg Met Val Gln Glu Arg Glu Lys Asp Ile Leu Thr Ala Ile
35 40 45
Ala Ala Asp Leu Cys Lys Ser Glu Phe Asn Val Tyr Ser Gln Glu
50 55 60
Val Ile Thr Val Leu Gly Glu Ile Asp Phe Met Leu Glu Asn Leu
65 70 75
Pro Glu Trp Val Thr Ala Lys Pro Val Lys Lys Asn Val Leu Thr
80 85 90
Met Leu Asp Glu Ala Tyr Ile Gln Pro Gln Pro Leu Gly Val Val
95 100 105
Leu Ile Ile Gly Ala Trp Asn Tyr Pro Phe Val Leu Thr Ile Gln
110 115 120
Pro Leu Ile Gly Ala Ile Ala Ala Gly Asn Ala Val Ile Ile Lys
125 130 135
Pro Ser Glu Leu Ser Glu Asn Thr Ala Lys Ile Leu Ala Lys Leu
140 145 150
Leu Pro Gln Tyr Leu Asp Gln Glu Phe Tyr Gly Glu Asn Ile Lys
155 160 165
Glu Ser Pro Asp Tyr Glu Arg Ile Ile Asn Leu Arg His Phe Lys
170 175 180
Arg Ile Leu Ser Leu Leu Glu Gly Gln Lys Ile Ala Phe Gly Gly
185 190 195
Glu Thr Asp Glu Ala Thr Arg Tyr Ile Ala Pro Thr Val Leu Thr
200 205 210
Asp Val Asp Pro Lys Thr Lys Val Met Gln Glu Glu Ile Phe Gly
215 220 225
Pro Ile Leu Pro Ile Val Pro Val Lys Asn Val Asp Glu Ala Ile

```
                230                     235                     240
Asn Phe Ile Asn Glu Arg Glu Lys Pro Leu Ala Leu Tyr Val Phe
                245                     250                     255
Ser His Asn His Lys Leu Ile Lys Arg Met Ile Asp Glu Thr Ser
                260                     265                     270
Ser Gly Gly Val Thr Gly Asn Asp Val Ile Met His Phe Thr Leu
                275                     280                     285
Asn Ser Phe Pro Phe Gly Gly Val Gly Ser Ser Gly Met Gly Ala
                290                     295                     300
Tyr His Gly Lys His Ser Phe Asp Thr Phe Ser His Gln Arg Pro
                305                     310                     315
Cys Leu Leu Lys Ser Leu Lys Arg Glu Gly Ala Asn Lys Leu Arg
                320                     325                     330
Tyr Pro Pro Asn Ser Gln Ser Lys Val Asp Trp Gly Lys Phe Phe
                335                     340                     345
Leu Leu Lys Arg Phe Asn Lys Glu Lys Leu Gly Leu Leu Leu Leu
                350                     355                     360
Thr Phe Leu Gly Ile Val Ala Ala Val Leu Val Lys Ala Glu Tyr
                365                     370                     375
Tyr


<210> 22
<211> 248
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512564CD1

<400> 22
Met Val Glu Leu Met Phe Pro Leu Leu Leu Leu Leu Leu Pro Phe
  1               5                      10                  15
Leu Leu Tyr Met Ala Ala Pro Gln Ile Arg Lys Met Leu Ser Ser
                 20                      25                  30
Gly Val Cys Thr Ser Thr Val Gln Leu Pro Gly Lys Val Val Val
                 35                      40                  45
Val Thr Gly Ala Asn Thr Gly Ile Gly Lys Glu Thr Ala Lys Glu
                 50                      55                  60
Leu Ala Gln Arg Gly Ala Arg Val Tyr Leu Ala Cys Arg Asp Val
                 65                      70                  75
Glu Lys Gly Glu Leu Val Ala Lys Glu Ile Gln Thr Thr Thr Gly
                 80                      85                  90
Asn Gln Gln Val Leu Val Arg Lys Leu Asp Leu Ser Asp Thr Lys
                 95                     100                 105
Ser Ile Arg Ala Phe Ala Lys Gly Phe Leu Ala Glu Glu Lys His
                110                     115                 120
Leu His Val Leu Ile Asn Asn Ala Gly Val Met Met Cys Pro Tyr
                125                     130                 135
Ser Lys Thr Ala Asp Gly Phe Glu Met His Ile Gly Val Asn His
                140                     145                 150
Leu Gly Ser Gly Val Thr Thr Tyr Ser Val His Pro Gly Thr Val
                155                     160                 165
Gln Ser Glu Leu Val Arg His Ser Ser Phe Met Arg Trp Met Trp
                170                     175                 180
```

```
Trp Leu Phe Ser Phe Phe Ile Lys Thr Pro Gln Gln Gly Ala Gln
                185                 190                 195
Thr Ser Leu His Cys Ala Leu Thr Glu Gly Leu Glu Ile Leu Ser
                200                 205                 210
Gly Asn His Phe Ser Asp Cys His Val Ala Trp Val Ser Ala Gln
                215                 220                 225
Ala Arg Asn Glu Thr Ile Ala Arg Arg Leu Trp Asp Val Ser Cys
                230                 235                 240
Asp Leu Leu Gly Leu Pro Ile Asp
                245
```

<210> 23
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512646CD1

<400> 23

```
Met Ala Ala Ala Val Gly Arg Leu Leu Arg Ala Ser Val Ala Arg
 1               5                  10                  15
His Val Ser Ala Ile Pro Trp Gly Ile Ser Ala Thr Ala Ala Leu
                20                  25                  30
Arg Pro Ala Ala Cys Gly Arg Thr Ser Leu Thr Asn Leu Leu Cys
                35                  40                  45
Ser Gly Ser Ser Gln Ala Lys Leu Phe Ser Thr Ser Ser Ser Cys
                50                  55                  60
His Ala Pro Ala Val Thr Gln His Ala Pro Tyr Phe Lys Gly Thr
                65                  70                  75
Ala Val Val Asn Gly Glu Phe Lys Asp Leu Ser Leu Asp Asp Phe
                80                  85                  90
Lys Gly Lys Tyr Leu Val Leu Phe Phe Tyr Pro Leu Asp Phe Thr
                95                  100                 105
Phe Val Cys Pro Thr Glu Ile Val Ala Phe Ser Asp Lys Ala Asn
                110                 115                 120
Glu Phe His Asp Val Asn Cys Glu Val Val Ala Val Ser Val Asp
                125                 130                 135
Ser His Phe Ser His Leu Ala Trp Ile Asn Thr Pro Arg Lys Asn
                140                 145                 150
Gly Gly Leu Gly His Met Asn Ile Ala Leu Leu Ser Asp Leu Thr
                155                 160                 165
Lys Gln Ile Ser Arg Asp Tyr Gly Val Leu Leu Glu Gly Ser Gly
                170                 175                 180
Leu Ala Leu Arg Ser Ser Gln Val Gln Leu Leu Pro Lys Ser Thr
                185                 190                 195
Phe Arg Arg
```

<210> 24
<211> 270
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7512700CD1

<400> 24
Met Asp Leu Ser Ala Ala Ser His Arg Ile Pro Leu Ser Asp Gly
1 5 10 15
Asn Ser Ile Pro Ile Ile Gly Leu Gly Thr Tyr Ser Glu Pro Lys
20 25 30
Ser Leu Trp Ala Thr Asn His Val Pro Glu Met Val Arg Pro Thr
35 40 45
Leu Glu Arg Thr Leu Arg Val Leu Gln Leu Asp Tyr Val Asp Leu
50 55 60
Tyr Ile Ile Glu Val Pro Met Ala Phe Lys Pro Gly Asp Glu Ile
65 70 75
Tyr Pro Arg Asp Glu Asn Gly Lys Trp Leu Tyr His Lys Ser Asn
80 85 90
Leu Cys Ala Thr Trp Glu Ala Met Glu Ala Cys Lys Asp Ala Gly
95 100 105
Leu Val Lys Ser Leu Gly Val Ser Asn Phe Asn Arg Arg Gln Leu
110 115 120
Glu Leu Ile Leu Asn Lys Pro Gly Leu Lys His Lys Pro Val Ser
125 130 135
Asn Gln Val Glu Cys His Pro Tyr Phe Thr Gln Pro Lys Leu Leu
140 145 150
Lys Phe Cys Gln Gln His Asp Ile Val Ile Thr Ala Tyr Ser Pro
155 160 165
Leu Gly Thr Ser Arg Asn Pro Ile Trp Val Asn Val Ser Ser Pro
170 175 180
Pro Leu Leu Lys Asp Ala Leu Leu Asn Ser Leu Gly Lys Arg Tyr
185 190 195
Asn Lys Thr Ala Ala Gln Ile Val Leu Arg Phe Asn Ile Gln Arg
200 205 210
Gly Val Val Val Ile Pro Lys Ser Phe Asn Leu Glu Arg Ile Lys
215 220 225
Glu Asn Phe Gln Ile Phe Asp Phe Ser Leu Thr Glu Glu Glu Met
230 235 240
Lys Asp Ile Glu Ala Leu Asn Lys Asn Val Arg Phe Val Glu Leu
245 250 255
Leu Met Trp Arg Asp His Pro Glu Tyr Pro Phe His Asp Glu Tyr
260 265 270

<210> 25
<211> 306
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512707CD1

<400> 25
Met Ala Ala Arg Leu Val Ser Arg Cys Gly Ala Val Arg Ala Ala
1 5 10 15
Pro His Ser Gly Pro Leu Val Ser Trp Arg Arg Trp Ser Gly Ala
20 25 30

Ser Thr Asp Thr Val Tyr Asp Val Val Val Ser Gly Gly Gly Leu
35 40 45
Val Gly Ala Ala Met Ala Cys Ala Leu Gly Tyr Asp Ile His Phe
50 55 60
His Asp Lys Lys Ile Leu Leu Leu Glu Ala Gly Pro Lys Lys Val
65 70 75
Leu Glu Lys Leu Ser Glu Thr Tyr Ser Asn Arg Val Ser Ser Ile
80 85 90
Ser Pro Gly Ser Ala Thr Leu Leu Ser Ser Phe Gly Ala Trp Asp
95 100 105
His Ile Cys Asn Met Arg Tyr Arg Ala Phe Arg Arg Met Gln Val
110 115 120
Trp Asp Ala Cys Ser Glu Ala Leu Ile Met Phe Asp Lys Asp Asn
125 130 135
Leu Asp Asp Met Gly Tyr Ile Val Glu Asn Asp Val Ile Met His
140 145 150
Ala Leu Thr Lys Gln Leu Glu Ala Val Ser Asp Arg Val Thr Val
155 160 165
Leu Tyr Arg Ser Lys Ala Ile Arg Tyr Thr Trp Pro Cys Pro Phe
170 175 180
Pro Met Ala Asp Ser Ser Pro Trp Val His Ile Thr Leu Gly Asp
185 190 195
Gly Ser Thr Phe Gln Thr Lys Leu Leu Ile Gly Ala Asp Gly His
200 205 210
Asn Ser Gly Val Arg Gln Ala Val Gly Ile Gln Asn Val Ser Trp
215 220 225
Asn Tyr Asp Gln Ser Ala Val Val Ala Thr Leu His Leu Ser Glu
230 235 240
Ala Thr Glu Asn Asn Val Ala Trp Gln Arg Phe Leu Pro Ser Gly
245 250 255
Pro Ile Ala Leu Leu Pro Leu Ser Asp Thr Leu Ser Ser Leu Val
260 265 270
Trp Ser Thr Ser His Glu His Ala Ala Glu Leu Val Ser Met Asp
275 280 285
Glu Glu Lys Phe Val Asp Ala Val Asn Ser Ala Phe Glu Gln Ile
290 295 300
Met Ala Phe Ala Ser Lys
305

<210> 26
<211> 423
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512710CD1

<400> 26
Met Ala Ala Arg Leu Val Ser Arg Cys Gly Ala Val Arg Ala Ala
1 5 10 15
Pro His Ser Gly Pro Leu Val Ser Trp Arg Arg Trp Ser Gly Ala
20 25 30
Ser Thr Asp Thr Val Tyr Asp Val Val Val Ser Gly Gly Gly Leu
35 40 45
Val Gly Ala Ala Met Ala Cys Ala Leu Gly Phe Gly Ala Trp Asp

```
                 50                   55                   60
His Ile Cys Asn Met Arg Tyr Arg Ala Phe Arg Arg Met Gln Val
                 65                   70                   75
Trp Asp Ala Cys Ser Glu Ala Leu Ile Met Phe Asp Lys Asp Asn
                 80                   85                   90
Leu Asp Asp Met Gly Tyr Ile Val Glu Asn Asp Val Ile Met His
                 95                  100                  105
Ala Leu Thr Lys Gln Leu Glu Ala Val Ser Asp Arg Val Thr Val
                110                  115                  120
Leu Tyr Arg Ser Lys Ala Ile Arg Tyr Thr Trp Pro Cys Pro Phe
                125                  130                  135
Pro Met Ala Asp Ser Ser Pro Trp Val His Ile Thr Leu Gly Asp
                140                  145                  150
Gly Ser Thr Phe Gln Thr Lys Leu Leu Ile Gly Ala Asp Gly His
                155                  160                  165
Asn Ser Gly Val Arg Gln Ala Val Gly Ile Gln Asn Val Ser Trp
                170                  175                  180
Asn Tyr Asp Gln Ser Ala Val Val Ala Thr Leu His Leu Ser Glu
                185                  190                  195
Ala Thr Glu Asn Asn Val Ala Trp Gln Arg Phe Leu Pro Ser Gly
                200                  205                  210
Pro Ile Ala Leu Leu Pro Leu Ser Asp Thr Leu Ser Ser Leu Val
                215                  220                  225
Trp Ser Thr Ser His Glu His Ala Ala Glu Leu Val Ser Met Asp
                230                  235                  240
Glu Glu Lys Phe Val Asp Ala Val Asn Ser Ala Phe Trp Ser Asp
                245                  250                  255
Ala Asp His Thr Asp Phe Ile Asp Thr Ala Gly Ala Met Leu Gln
                260                  265                  270
Tyr Ala Val Ser Leu Leu Lys Pro Thr Lys Val Ser Ala Arg Gln
                275                  280                  285
Leu Pro Pro Ser Val Ala Arg Val Asp Ala Lys Ser Arg Val Leu
                290                  295                  300
Phe Pro Leu Gly Leu Gly His Ala Ala Glu Tyr Val Arg Pro Arg
                305                  310                  315
Val Ala Leu Ile Gly Asp Ala Ala His Arg Val His Pro Leu Ala
                320                  325                  330
Gly Gln Gly Val Asn Met Gly Phe Gly Asp Ile Ser Ser Leu Ala
                335                  340                  345
His His Leu Ser Thr Ala Ala Phe Asn Gly Lys Asp Leu Gly Ser
                350                  355                  360
Val Ser His Leu Thr Gly Tyr Glu Thr Glu Arg Gln Arg His Asn
                365                  370                  375
Thr Ala Leu Leu Ala Ala Thr Asp Leu Leu Lys Arg Leu Tyr Ser
                380                  385                  390
Thr Ser Ala Ser Pro Leu Val Leu Leu Arg Thr Trp Gly Leu Gln
                395                  400                  405
Ala Thr Asn Ala Val Ser Pro Leu Lys Glu Gln Ile Met Ala Phe
                410                  415                  420
Ala Ser Lys
```

<210> 27
<211> 88
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512884CD1

<400> 27
Met Glu Gln Thr Glu Val Leu Lys Pro Arg Thr Leu Ala Asp Leu
1 5 10 15
Ile Arg Ile Leu His Gln Leu Phe Ala Gly Asp Glu Val Asn Val
20 25 30
Glu Glu Val Gln Ala Ile Met Glu Ala Tyr Glu Ser Asp Pro Thr
35 40 45
Glu Trp Ala Met Tyr Ala Lys Phe Asp Gln Tyr Arg Tyr Thr Arg
50 55 60
Asn Leu Val Asp Gln Gly Asn Gly Lys Phe Asn Leu Met Ile Leu
65 70 75
Cys Trp Gly Glu Gly His Gly Arg Cys Tyr Arg Glu Ile
80 85

<210> 28
<211> 458
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512931CD1

<400> 28
Met Gly Arg Thr Val Val Val Leu Gly Gly Gly Ile Ser Gly Leu
1 5 10 15
Ala Ala Ser Tyr His Leu Ser Arg Ala Pro Cys Pro Pro Lys Val
20 25 30
Val Leu Val Glu Ser Ser Glu Arg Leu Gly Gly Trp Ile Arg Ser
35 40 45
Val Arg Gly Pro Asn Gly Ala Ile Phe Glu Leu Gly Pro Arg Gly
50 55 60
Ile Arg Pro Ala Gly Ala Leu Gly Ala Arg Thr Leu Leu Leu Val
65 70 75
Ser Glu Leu Gly Leu Asp Ser Glu Val Leu Pro Val Arg Gly Asp
80 85 90
His Pro Ala Ala Gln Asn Arg Phe Leu Tyr Val Gly Gly Ala Leu
95 100 105
His Ala Leu Pro Thr Gly Leu Arg Glu Leu Thr Lys Pro Arg Gly
110 115 120
Lys Glu Pro Asp Glu Thr Val His Ser Phe Ala Gln Arg Arg Leu
125 130 135
Gly Pro Glu Val Ala Ser Leu Ala Met Asp Ser Leu Cys Arg Gly
140 145 150
Val Phe Ala Gly Asn Ser Arg Glu Leu Ser Ile Arg Ser Cys Phe
155 160 165
Pro Ser Leu Phe Gln Ala Glu Gln Thr His Arg Ser Ile Leu Leu
170 175 180
Gly Leu Leu Leu Gly Ala Gly Arg Thr Pro Gln Pro Asp Ser Ala
185 190 195
Leu Ile Arg Gln Ala Leu Ala Glu Arg Trp Ser Gln Trp Ser Leu
200 205 210

Arg Gly Gly Leu Glu Met Leu Pro Gln Ala Leu Glu Thr His Leu
215 220 225
Thr Ser Arg Gly Val Ser Val Leu Arg Gly Gln Pro Val Cys Gly
230 235 240
Leu Ser Leu Gln Ala Glu Gly Arg Trp Lys Val Ser Leu Arg Asp
245 250 255
Ser Ser Leu Glu Ala Asp His Val Ile Ser Ala Ile Pro Ala Ser
260 265 270
Val Leu Ser Glu Leu Leu Pro Ala Glu Ala Ala Pro Leu Ala Arg
275 280 285
Ala Leu Ser Ala Ile Thr Ala Val Ser Val Ala Val Val Asn Leu
290 295 300
Gln Tyr Gln Gly Ala His Leu Pro Val Gln Gly Phe Gly His Leu
305 310 315
Val Pro Ser Ser Glu Asp Pro Gly Val Leu Gly Ile Val Tyr Asp
320 325 330
Ser Val Ala Phe Pro Glu Gln Asp Gly Ser Pro Pro Gly Leu Arg
335 340 345
Val Thr Val Met Leu Gly Gly Ser Trp Leu Gln Thr Leu Glu Ala
350 355 360
Ser Gly Cys Val Leu Ser Gln Glu Leu Phe Gln Gln Arg Ala Gln
365 370 375
Glu Ala Ala Ala Thr Gln Leu Gly Leu Lys Glu Met Pro Ser His
380 385 390
Cys Leu Val His Leu His Lys Asn Cys Ile Pro Gln Tyr Thr Leu
395 400 405
Gly His Trp Gln Lys Leu Glu Ser Ala Arg Gln Phe Leu Thr Ala
410 415 420
His Arg Leu Pro Leu Thr Leu Ala Gly Ala Ser Tyr Glu Gly Val
425 430 435
Ala Val Asn Asp Cys Ile Glu Ser Gly Arg Gln Ala Ala Val Ser
440 445 450
Val Leu Gly Thr Glu Pro Asn Ser
455

<210> 29
<211> 185
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512933CD1

<400> 29
Met Gly Arg Thr Val Val Val Leu Gly Gly Gly Ile Ser Gly Leu
1 5 10 15
Ala Ala Ser Tyr His Leu Ser Arg Ala Pro Cys Pro Pro Lys Val
20 25 30
Val Leu Val Glu Ser Ser Glu Arg Leu Gly Gly Trp Ile Arg Ser
35 40 45
Val Arg Gly Pro Asn Gly Ala Ile Phe Glu Leu Gly Pro Arg Gly
50 55 60
Ile Arg Pro Ala Gly Ala Leu Gly Ala Arg Thr Leu Leu Leu Val
65 70 75
Met Leu Gly Gly Ser Trp Leu Gln Thr Leu Glu Ala Ser Gly Cys

80 85 90
Val Leu Ser Gln Glu Leu Phe Gln Gln Arg Ala Gln Glu Ala Ala
95 100 105
Ala Thr Gln Leu Gly Leu Lys Glu Met Pro Ser His Cys Leu Val
110 115 120
His Leu His Lys Asn Cys Ile Pro Gln Tyr Thr Leu Gly His Trp
125 130 135
Gln Lys Leu Glu Ser Ala Arg Gln Phe Leu Thr Ala His Arg Leu
140 145 150
Pro Leu Thr Leu Ala Gly Ala Ser Tyr Glu Gly Val Ala Val Asn
155 160 165
Asp Cys Ile Glu Ser Gly Arg Gln Ala Ala Val Ser Val Leu Gly
170 175 180
Thr Glu Pro Asn Ser
185

<210> 30
<211> 140
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512942CD1

<400> 30
Met Gly Arg Thr Val Val Val Leu Gly Gly Gly Ile Ser Gly Leu
1 5 10 15
Ala Ala Ser Tyr His Leu Ser Arg Ala Pro Cys Pro Pro Lys Val
20 25 30
Met Leu Gly Gly Ser Trp Leu Gln Thr Leu Glu Ala Ser Gly Cys
35 40 45
Val Leu Ser Gln Glu Leu Phe Gln Gln Arg Ala Gln Glu Ala Ala
50 55 60
Ala Thr Gln Leu Gly Leu Lys Glu Met Pro Ser His Cys Leu Val
65 70 75
His Leu His Lys Asn Cys Ile Pro Gln Tyr Thr Leu Gly His Trp
80 85 90
Gln Lys Leu Glu Ser Ala Arg Gln Phe Leu Thr Ala His Arg Leu
95 100 105
Pro Leu Thr Leu Ala Gly Ala Ser Tyr Glu Gly Val Ala Val Asn
110 115 120
Asp Cys Ile Glu Ser Gly Arg Gln Ala Ala Val Ser Val Leu Gly
125 130 135
Thr Glu Pro Asn Ser
140

<210> 31
<211> 281
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7513736CD1

<400> 31
Met Ser Leu Val Cys Leu Thr Asp Phe Gln Ala His Ala Arg Glu
1 5 10 15
Gln Leu Ser Lys Ser Thr Arg Asp Phe Ile Glu Gly Gly Ala Asp
20 25 30
Asp Ser Ile Thr Arg Asp Asp Asn Ile Ala Ala Phe Lys Arg Ile
35 40 45
Arg Leu Arg Pro Arg Tyr Leu Arg Asp Val Ser Glu Val Asp Thr
50 55 60
Arg Thr Thr Ile Gln Gly Glu Glu Ile Ser Ala Pro Ile Cys Ile
65 70 75
Ala Pro Thr Gly Phe His Cys Leu Val Trp Pro Asp Gly Glu Met
80 85 90
Ser Thr Ala Arg Ala Ala Gln Ala Ala Gly Ile Cys Tyr Ile Thr
95 100 105
Ser Thr Phe Ala Ser Cys Ser Leu Glu Asp Ile Val Ile Ala Ala
110 115 120
Pro Glu Gly Leu Arg Trp Phe Gln Leu Tyr Val His Pro Asp Leu
125 130 135
Gln Leu Asn Lys Gln Leu Ile Gln Arg Val Glu Ser Leu Gly Phe
140 145 150
Lys Ala Leu Val Ile Thr Leu Asp Thr Pro Val Cys Gly Asn Arg
155 160 165
Arg His Asp Ile Arg Asn Gln Leu Arg Arg Asn Leu Thr Leu Thr
170 175 180
Asp Leu Gln Ser Pro Lys Lys Ile Asp Ala Leu Thr Glu Val Val
185 190 195
Ala Ala Val Lys Gly Lys Ile Glu Val Tyr Leu Asp Gly Gly Val
200 205 210
Arg Thr Gly Asn Asp Val Leu Lys Ala Leu Ala Leu Gly Ala Lys
215 220 225
Cys Ile Phe Leu Gly Arg Pro Ile Leu Trp Gly Leu Ala Cys Lys
230 235 240
Gly Glu His Gly Val Lys Glu Val Leu Asn Ile Leu Thr Asn Glu
245 250 255
Phe His Thr Ser Met Ala Leu Thr Gly Cys Arg Ser Val Ala Glu
260 265 270
Ile Asn Arg Asn Leu Val Gln Phe Ser Arg Leu
275 280

<210> 32
<211> 463
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512779CD1

<400> 32
Met Ala Asp Ser Glu Ala Leu Pro Ser Leu Ala Gly Asp Pro Val
1 5 10 15
Ala Val Glu Ala Leu Leu Arg Ala Val Phe Gly Val Val Val Asp
20 25 30
Glu Ala Ile Gln Lys Gly Thr Ser Val Ser Gln Lys Val Cys Glu
35 40 45

Trp Lys Glu Pro Glu Glu Leu Lys Gln Leu Leu Asp Leu Glu Leu
50 55 60
Arg Ser Gln Gly Glu Ser Gln Lys Gln Ile Leu Glu Arg Cys Arg
65 70 75
Ala Val Ile Arg Tyr Ser Val Lys Thr Gly His Pro Arg Phe Phe
80 85 90
Asn Gln Leu Phe Ser Gly Leu Asp Pro His Ala Leu Ala Gly Arg
95 100 105
Ile Ile Thr Glu Ser Leu Asn Thr Ser Gln Tyr Thr Tyr Glu Ile
110 115 120
Ala Pro Val Phe Val Leu Met Glu Glu Glu Val Leu Arg Lys Leu
125 130 135
Arg Ala Leu Val Gly Trp Ser Ser Gly Asp Gly Ile Phe Cys Pro
140 145 150
Gly Gly Ser Ile Ser Asn Met Tyr Ala Val Asn Leu Ala Arg Tyr
155 160 165
Gln Arg Tyr Pro Asp Cys Lys Gln Arg Gly Leu Arg Thr Leu Pro
170 175 180
Pro Leu Ala Leu Phe Thr Ser Lys Glu Cys His Tyr Ser Ile Gln
185 190 195
Lys Gly Ala Ala Phe Leu Gly Leu Gly Thr Asp Ser Val Arg Val
200 205 210
Val Lys Ala Asp Glu Arg Gly Lys Met Val Pro Glu Asp Leu Glu
215 220 225
Arg Gln Ile Gly Met Ala Glu Ala Glu Gly Ala Val Pro Phe Leu
230 235 240
Val Ser Ala Thr Ser Gly Thr Thr Val Leu Gly Ala Phe Asp Pro
245 250 255
Leu Glu Ala Ile Ala Asp Val Cys Gln Arg His Gly Leu Trp Leu
260 265 270
His Val Asp Ala Ala Trp Gly Gly Ser Val Leu Leu Ser Gln Thr
275 280 285
His Arg His Leu Leu Asp Gly Ile Gln Arg Ala Asp Ser Val Ala
290 295 300
Trp Asn Pro His Lys Leu Leu Ala Ala Gly Leu Gln Cys Ser Ala
305 310 315
Leu Leu Leu Gln Asp Thr Ser Asn Leu Leu Lys Arg Cys His Gly
320 325 330
Ser Gln Ala Ser Tyr Leu Phe Gln Gln Asp Lys Phe Tyr Asp Val
335 340 345
Ala Leu Asp Thr Gly Asp Lys Val Val Gln Cys Gly Arg Arg Val
350 355 360
Asp Cys Leu Lys Leu Trp Leu Met Trp Lys Ala Gln Gly Asp Gln
365 370 375
Gly Leu Glu Arg Arg Ile Asp Gln Ala Phe Val Leu Ala Arg Tyr
380 385 390
Leu Val Glu Glu Met Lys Lys Arg Glu Gly Phe Glu Leu Val Met
395 400 405
Glu Pro Ala Arg Glu Ala Gly Glu Ser Arg Leu Pro Arg Lys Ala
410 415 420
Val Lys Gly Gly Pro Arg Ala Gln Gly Ala His Gly Glu Gly Gly
425 430 435
Leu His Asp Asp Trp Leu Pro Ala Pro Arg Asp Pro Gly Gln Leu
440 445 450
Leu Pro Cys Gly Cys Gly Gln Leu Cys Thr Asp Leu Cys
455 460

<210> 33
<211> 149
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512877CD1

<400> 33
Met Cys Arg Thr Ala Leu Gly Val Thr Thr Leu Gln Asp Leu Gly
1 5 10 15
Val Cys Trp Arg Lys Glu His Asp Pro Ala Glu Met Leu Arg Val
20 25 30
Gly Arg Ala Glu Leu Gly Gly Trp Lys Pro Ser Trp Glu Ala Ile
35 40 45
Val Lys Glu Val Ala Ala Ala Thr Gly His Thr Val Val Leu Val
50 55 60
Asp Gln Thr Glu Asp Ile Leu Ala Lys Ser Lys Lys Gly Ile Glu
65 70 75
Glu Ser Leu Arg Lys Val Ala Lys Lys Lys Phe Ala Glu Asn Pro
80 85 90
Lys Ala Gly Asp Glu Phe Val Glu Lys Thr Leu Ser Thr Ile Ala
95 100 105
Thr Ser Thr Asp Ala Ala Ser Val Val His Ser Thr Asp Leu Val
110 115 120
Val Glu Ala Ile Val Glu Asn Leu Lys Val Lys Asn Glu Leu Phe
125 130 135
Lys Arg Leu Asp Lys Phe Ala Ala Glu Ser Leu Lys His Gln
140 145

<210> 34
<211> 203
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512782CD1

<400> 34
Met Asp Ala Glu Gly Leu Ala Leu Leu Leu Pro Pro Val Thr Leu
1 5 10 15
Ala Ala Leu Val Asp Ser Trp Leu Arg Glu Asp Cys Pro Gly Leu
20 25 30
Asn Tyr Ala Ala Leu Val Ser Gly Ala Gly Pro Ser Gln Ala Ala
35 40 45
Leu Trp Ala Lys Ser Pro Gly Val Leu Ala Gly Gln Pro Phe Phe
50 55 60
Asp Ala Ile Phe Thr Gln Leu Asn Cys Gln Val Ser Trp Phe Leu
65 70 75
Pro Glu Gly Ser Lys Leu Val Pro Val Ala Arg Val Ala Glu Ala
80 85 90
Val Arg Ala Ala Arg Gln Ala Ala Asp Phe Ala Leu Lys Val Glu
95 100 105
Val Glu Cys Ser Ser Leu Gln Glu Ala Val Gln Ala Ala Glu Ala

110 115 120
Gly Ala Asp Leu Val Leu Leu Asp Asn Phe Lys Pro Glu Glu Leu
125 130 135
His Pro Thr Ala Thr Val Leu Lys Ala Gln Phe Pro Ser Val Ala
140 145 150
Val Glu Ala Ser Gly Gly Ile Thr Leu Asp Asn Leu Pro Gln Phe
155 160 165
Cys Gly Pro His Ile Asp Val Ile Ser Met Gly Met Leu Thr Gln
170 175 180
Ala Ala Pro Ala Leu Asp Phe Ser Leu Lys Leu Phe Ala Lys Glu
185 190 195
Val Ala Pro Val Pro Lys Ile His
200

<210> 35
<211> 166
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512784CD1

<400> 35
Met Asp Ala Glu Gly Leu Ala Leu Leu Leu Pro Pro Val Thr Leu
1 5 10 15
Ala Ala Leu Val Asp Ser Trp Leu Arg Glu Asp Cys Pro Gly Leu
20 25 30
Asn Tyr Ala Ala Leu Val Ser Gly Ala Gly Pro Ser Gln Ala Ala
35 40 45
Leu Trp Ala Lys Ser Pro Gly Ala Val Arg Ala Ala Arg Gln Ala
50 55 60
Ala Asp Phe Ala Leu Lys Val Glu Val Glu Cys Ser Ser Leu Gln
65 70 75
Glu Ala Val Gln Ala Ala Glu Ala Gly Ala Asp Leu Val Leu Leu
80 85 90
Asp Asn Phe Lys Pro Glu Glu Leu His Pro Thr Ala Thr Val Leu
95 100 105
Lys Ala Gln Phe Pro Ser Val Ala Val Glu Ala Ser Gly Gly Ile
110 115 120
Thr Leu Asp Asn Leu Pro Gln Phe Cys Gly Pro His Ile Asp Val
125 130 135
Ile Ser Met Gly Met Leu Thr Gln Ala Ala Pro Ala Leu Asp Phe
140 145 150
Ser Leu Lys Leu Phe Ala Lys Glu Val Ala Pro Val Pro Lys Ile
155 160 165
His

<210> 36
<211> 48
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7512794CD1

<400> 36
Met Glu Ala Ala His Phe Phe Glu Gly Thr Glu Lys Leu Leu Glu
1 5 10 15
Val Trp Phe Ser Arg Gln Gln Pro Asp Ala Asn Gln Gly Ser Gly
20 25 30
Asp Leu Arg Thr Ile Pro Arg Trp Ser Ser Ile Leu Tyr Gly Thr
35 40 45
Tyr Glu Phe

<210> 37
<211> 244
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512886CD1

<400> 37
Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu Ala Val Leu Ala
1 5 10 15
Trp Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg Met Lys Ser
20 25 30
Arg Glu Gln Gly Gly Arg Leu Gly Ala Glu Ser Arg Thr Leu Leu
35 40 45
Val Ile Ala His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro Thr
50 55 60
Val Leu Gly Leu Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys
65 70 75
Phe Ser Ala Gly Asn Tyr Tyr Asn Gln Gly Glu Thr Arg Lys Lys
80 85 90
Glu Leu Leu Gln Ser Cys Asp Val Leu Gly Ile Pro Leu Ser Ser
95 100 105
Val Met Ile Ile Asp Asn Arg Asp Phe Pro Asp Asp Pro Gly Met
110 115 120
Gln Trp Asp Thr Glu His Val Ala Arg Val Leu Leu Gln His Ile
125 130 135
Glu Val Asn Gly Ile Asn Leu Val Val Thr Phe Asp Ala Gly Gly
140 145 150
Val Ser Gly His Ser Asn His Ile Ala Leu Tyr Ala Ala Val Arg
155 160 165
Val Leu Cys Ala His Ala Ser Val Cys Glu Cys Ala Ala Gln Val
170 175 180
His Leu Pro Ser Gly Ser Ala Leu Val Ser Ala Ser Tyr Ala Gly
185 190 195
Cys Pro Leu Arg Ala Gln Gln Gln Arg Ser Gly Thr Gly Gln Glu
200 205 210
Ser His Val Leu Pro Pro Gln Pro Ala Pro Leu Val Pro Pro Pro
215 220 225
Leu His Tyr Leu Leu Pro Val His Glu Asn Gln Leu Thr Glu Leu
230 235 240
Pro Leu Lys Pro

<210> 38
<211> 115
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512929CD1

<400> 38
Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg
1   5   10   15
Arg Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val
                20   25   30
Ala Ser Val Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu
                35   40   45
Asn Gly Glu Leu Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu
                50   55   60
Pro Leu Arg Glu Lys Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln
                65   70   75
Lys Tyr Pro Pro Val Lys Phe Leu Ser Glu Lys Asp Arg Lys Arg
                80   85   90
Ile Leu Leu Thr Arg Tyr Thr Ile Trp His Leu Gln Pro Pro Leu
                95   100   105
Gln Thr Thr Cys Ile Ile Leu Ser Arg His
                110   115

<210> 39
<211> 1798
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511289CB1

<400> 39
gctctgcaga gtggtggccg gggccagggc cggggtgccc tccctcccac cttctcccgc 60
catgagccag ggaagtccgg gggactgggc ccccctagat cccacccccg gacccccagc 120
atcccccaac cccttcgtgc atgagttaca tctctctcgc ctccagaggg ttaagttctg 180
cctcctgggg gcattgctgg cccccatccg agtgcttctg gcctttatcg tcctctttct 240
cctctggccc tttgcctggc ttcaagtggc cggtcttagt gaggagcagc ttcaggagcc 300
aattacagga tggaggaaga ctgtgtgcca caacggggtg ctaggcctga gccgcctgct 360
gtttttcctg ctgggcttcc tccggattcg cgttcgtggc cagcgagcct ctcgccttca 420
agcccctgtc cttgttgctg ccccacactc cactttcttt gaccccattg ttctgctgcc 480
ctgtgacctg cccaaagttg tgtcccgagc tgagaacctt tccgttcctg tcattggagc 540
ccttcttcga ttcaaccaag ccatcctggt atcccggcat gacccggctt ctcgacgcag 600
agtggtggag gaggtccgaa ggcgggccac ctcaggaggc aagtggccgc aggtgctatt 660
ctttcctgag ggcacctgtt ccaacaagaa ggctttgctt aagttcaaac caggacacca 720
ccagctgggc atggaggggt cctggagtac tcaaagtcct ctggctcaca gcctctcagc 780
cctgcagcat tgtggatgtg gagttccttc ctgtgtatca ccccagccct gaggagagca 840
gggaccccac cctctatgcc aacaatgttc agagggtcat ggcacaggct ctgggcattc 900
cagccaccga atgtgagttt gtagggagct tacctgtgat tgtggtgggc cggctgaagg 960
tggcgttgga accacagctc tgggaactgg gaaaagtgct tcggaaggct gggctgtccg 1020
ctggctatgt ggacgctggg gcagagccag gccggagtcg aatgatcagc caggaagagt 1080
ttgccaggca gctacagctc tctgatcctc agacggtggc tggtgccttt ggctacttcc 1140

agcaggatac caagggtttg gtggacttcc gagatgtggc ccttgcacta gcagctctgg 1200
atgggggcag gagcctggaa gagctaactc gtctggcctt tgagctcttt gctgaagagc 1260
aagcagaggg tcccaaccgc ctgctgtaca aagacggctt cagcaccatc ctgcacctgc 1320
tgctgggttc accccaccct gctgccacag ctttgcatgc tgagctgtgc caggcaggat 1380
ccagccaagg cctctccctc tgtcagttcc agaacttctc cctccatgac ccactctatg 1440
ggaaactctt cagcacctac ctgcgccccc cacacacctc tcgaggcacc tcccagacac 1500
caaatgcctc atccccaggc aaccccactg ctctggccaa tgggactgtg caagcaccca 1560
agcagaaggg agactgagtg cctcagcctc tcacccccctc ctcctcaggg cagcgctagg 1620
ggcctcccct atgcctcagc cccatctctg ctcctgtttg aattttgtta ttgttgtttg 1680
gttgttgttt ttttaagttg attttaattt tttgtttggt tgattttttt gtaaaaaact 1740
attttatata taaatataaa tctatatcta tatctattaa aaaaaatgaa gtccaaaa 1798

<210> 40
<211> 929
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511056CB1

<400> 40
gcgccatcag gcccgagata gcggcgaggt ccgctttcag tgtatggttt tccctgccaa 60
acggttctgc ttggtgccat ccatggaggg cgtgcgctgg gccttttcct gcggcacttg 120
gctgccgagc cgagccgaat ggctgctggc agtgcgatcg attcagcccg aggagaagga 180
gcgcattggc cagttcgtct ttgcccggga cgctaaggca gccatggctg gtcgtctgat 240
gataaggaaa ttagttgcag agaaattgaa tatcccttgg aatcatattc gtttgcaaag 300
aactgcaaaa ggaaaaccag ttcttgcaaa ggactcatcg aatccttacc cgaatttcaa 360
ctttaacatc tctcatcaag gagactatgc agtgcttgct gctgaacctg agctgcaagt 420
tggaattgat ataatgaaga ctagttttcc aggcacttaa ggaaagcttc ataaaagcca 480
ttggtgttgg actaggattt gaattgcagc ggcttgaatt tgatctatct ccattaaact 540
tggatatagg ccaagtttat aaagaaacac gtttattcct ggatggagag gaagaaaaag 600
aatgggcatt tgaggaaagc aaaatagatg agcaccattt tgttgcagtt gctcttagga 660
aacccgatgg atctagacat caggatgttc catctcagga tgattccaaa ccaacccaga 720
ggcaatttac tattctcaac tttaatgatt taatgtcatc tgccgttccc atgacacctg 780
aagatccttc attttgggac tgtttttgct tcacagaaga aattccaata cgaaatggta 840
caaagtcatg atgattccct gagtaacaaa gggaaatgaa aactgtttgt gatcttccgt 900
attcactgaa aaataaatgc ttgtttagt 929

<210> 41
<211> 2163
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511567CB1

<400> 41
cgaattctgc tggaggcagc gccatatcct ggaggtgaag gtaccacctc atggagaccc 60
ccgcggccgc cgccccccgct gggagcttat tccccctcctt cctgctcctg gcctgcggga 120
cgctggtggc cgcccttgctg ggcgccgctc accgcctggg gctcttctat cagctgctgc 180
acaaggtgga caaggcaagc gtccggcatg gcggagagaa cgtggccgct gtgctgaggg 240
cccatggtgt gcggttcatc ttcacgctgg tcggtgggca catttccccg ctgctggtgg 300
cctgtgagaa actgggcatc cgtgtggtgg acacacgcca tgaggtcacg gccgtctttg 360
ctgctgatgc tatggcccgc ctgtccggga cggtgggcgt ggcggcagtg acagcaggcc 420

ctggcctcac caacacggtg actgcggtga agaatgctca gatggctcag tccccaatcc 480
tgcttctggg tggggctgcc agcactctgc tgcagaaccg gggtgcgctc caggctgttg 540
atcagctgtc ccttttccgg ccactctgta agttttgtgt gtctgtgcgg agggtgcggg 600
acattgtgcc caccctgagg gccgcgatgg ctgccgccca gtcgggcacc ccaggtccgg 660
tgtttgtgga gctgcccgtt gacgtgcttt acccctactt catggtccag aaggagatgg 720
tgccagccaa gccacccaag ggcctcgtgg gccgagtggt ctcctggtat ttagagaatt 780
acctggccaa cctctttgca ggagcctggg agcctcagcc cgagggaccg ctgcccctgg 840
acatccccca ggcttccccg cagcaggttc agcgctgtgt ggagatcctg agccgggcca 900
agaggcctct gatggtgctg gggagtcagg ccctgctcac ccccaacgtct gccgacaagc 960
ttcgggctgc cgtggagacc ttgggtgttc cctgcttcct tggagggatg gcacgggggc 1020
tgttaggccg caaccacccc ctccacatcc gggagaaccg cagtgcggcc ctgaagaagg 1080
cggatgtcat tgtcctagca ggaactgtgt gtgacttccg cctatcctat ggccgtgtcc 1140
tcagccacag cagcaagatc atcatcgtca atcgtaatcg ggaagagatg ttgctcaact 1200
cagacatctt ctggaagccc caggaggctg tgcagggaga tgtgggttcc ttcgtgctga 1260
agttagtgga gggccttcag ggccagacct gggccccaga ctgggtggag gagctgcggg 1320
aagccgaccg gcagaaggag cagacctttc gggagaaggc agcgatgcct gtggcccagc 1380
acctgaaccc agtgcaggtg ctgcagctgg tggaggaaac gctacctgac aactcaattc 1440
tggtggtgga tggtggggac ttcgtgggca ctgctgccca tctggtacag ccccgcggcc 1500
ccctgcgctg gcttgatcct ggggcctttg ggactctggg agttggtgca ggatttgcac 1560
ttggggccaa gctgtgccgg ccagatgctg aggtctggtg cctgtttggg gacggagctt 1620
ttggctacag cctcatcgaa tttgatacat tcgtcagaca caagattatc acaaggcagc 1680
catgggtctg gggggcccggg gcttgctgct ctcacgggag aacgaggatc aggtggtcaa 1740
ggtgctgcac gatgcccagc agcagtgccg agacggccac ccggttgtgg tcaacatcct 1800
cattgggagg acggacttcc gcgatggctc cattgctgta tagggccttg tgggtcagga 1860
cgcttggctg ccttcctgcc cctggactgt gtccagctgg tttggagtct catcattgct 1920
tgccctgggc ctatcccatg aggcccagtg actccacaac ctccctgagg aggggatgga 1980
ggggtcaaat ctcagagagg ctctcttgtg agctcttgga ccctcctctt cacggaccca 2040
actgtgccgt ttgtcccctc tcttatggag actgaataaa ccacctctgg cccatgtagg 2100
cccaagtact cagaaaaaaa aaaaaaaaaa aaaaaaaagg gcgacctcgc gatctagaac 2160
tag 2163

<210> 42
<211> 1260
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511651CB1

<400> 42
ggtgcgcatg cgcgagcgcc tagctgcgcg caccacccac gagaccccag tctgaggtgc 60
ggggtcctgg ggcccggcgc gggtggccgc cgcggcccct cgggctgcgt ggggaggggg 120
cttccgcccc tgttgtcatt gctcctgcag ccttttcgct gggactgcgc gacaccgccc 180
cccgaccggg tgcccgctgt gtgccaggcc gggtgctggg cacggtcccg cgagtgccct 240
ataaggactg ccaggcaata atgaaggttc ttttactgaa ggatgcgaag gaagatgact 300
gtggccagga tccgtatatc aggctttctc atcctgaaga ttacgggggga ctcattttta 360
ccagccccag agcagtggaa gcagcagagt tatgtttgga gcaaaacaat aaaactgaag 420
tctgggaaag gtctctgaaa gaaaaatgga atgccaagtc agtgtatgtg gttggaaatg 480
ctactgcttc tctagtgagt aaaattggcc tggatacaga aggagaaacc tgtggaaatg 540
cagaaaagct tgcagaatat atttgttcca gggagtcctc agcactgcct cttctatttc 600
cctgtggaaa cctcaaaaga gaaatcctgc caaaagcgct caaggacaaa gggattgcca 660
tggaaagcat aactgtgtat cagacagttg cacacccagg aatccaaggg aacctgaaca 720
gctactattc ccagcagggg gttccagcca gcatcacatt ttttagtccc tctggcctca 780
catacagtct caagcacatt caggagttat ctggtgacaa tatcgatcaa attaagtttg 840
cagccatcgg ccccactacg gctcgcgcgc tggccgccca gggccttcct gtaagctgca 900

ctgcagagag ccccacgcca caagccctgg ccactggcat caggaaggct ctccagcccc 960
atggctgctg ctgagtcagc cacctagcgc tggccccatg cagcctccct gggctgggct 1020
ggctctggat ggagccaggc atcggcaagg gctctcggga gctgctgccg tcagactcct 1080
gcctcaagcc tgagtggaag cacctgagga ccggggatcg ggacctgacc tggggctggc 1140
ctcaggccca cgtgcacgtg actgccctct gtggaagcca gcttaaaccc tagccctgtg 1200
agagcttcct gtgcccagca ggaaggaagt caaataaacc acactgacta cccccaaaaa 1260

<210> 43
<211> 772
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511881CB1

<400> 43
gggagagcac cgaggctgct cttcctctcc tgggccgtcc tctgagcagc agacggggct 60
aagcgttccc cagctcgcct tcacacacag cccgtgccac cacaccgacg gtaccatgaa 120
ggacgaagta gctctactgg ctgctgtcac cctcctggga gtcctgctgc aagcctactt 180
ctccctgcag gtgatctcgg cgcgcagggc cttccgcgtg tcgccgccgc tcaccaccgg 240
cccacccgag ttcgagcgcg tctaccgagc ccaggtgaac tgcagcgagt acttcccgct 300
gttcctcgcc acgctctggg tcgccggcat cttctttcat gaaggtcggg gtgtggggca 360
ggggcgcacg cgctggaccc ccgggacccg cgcagggcgc tcaccaggcc cgtgcgtacc 420
tctcgcaggg gcggcggccc tgtgcggcct ggtctacctg ttcgcgcgcc tccgctactt 480
ccagggctac gcgcgctccg cgcagctcag gctggcaccg ctgtacgcga gcgcgcgcgc 540
cctctggctg ctggtggcgc tggctgcgct cggcctgctc gcccacttcc tcccggccgc 600
gctgcgcgcc gcgctcctcg gacggctccg gacgctgctg ccgtgggcct gagaccaagg 660
cccccgggcc gacggagccg ggaaagaaga gccggagcct ccagctgccc cggggagggg 720
cgctcgcttc cgcatcctag tctctatcat taaagttcta gtgaccgaga cc 772

<210> 44
<211> 911
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512181CB1

<400> 44
gccgctgggc accccgagac catggggaag ctcgtggcgc tggtcctgct gggggtcggc 60
ctgtccttag tcggggagat gttcctggcg tttagagaaa gggtgaatgc ctctcgagaa 120
gtggagccag tagaacctga aaactgccac cttattgagg aacttgaaag tggctctgaa 180
gatattgata tacttcctag tgggctggct tttatctcca gtggattaaa atatccaggc 240
atgccaaact ttgcgccaga tgaaccagga aaaatcttct tgatggatct gaatgaacaa 300
aacccaaggg cacaagcgct agaaatcagt ggtggatttg acaaagaatt atttaatcca 360
catgggatca gtattttcat cgacaaagac aatactgtgt atctttatgt tgtgaatcat 420
ccccacatga agtccactgt ggagatattt aaatttgagg aacaacaacg ttctctggta 480
tacctgaaaa ctataaaaca tgaacttctc aaaagtgtga atgacattgt ggttcttgga 540
ccagaacagt tctatgccac cagagaccac tattttacca actccctcct gtcatttttt 600
gagatgatct tggatcttcg ctggacttat gttcttttct acagcccaag ggaggttaaa 660
gtggtggcca aaggattttg tagtgccaat gggatcacag tctcagcaga ccagaagtac 720
ttcgcatcca gaatgttttg tctgagaagc ccagggtgag caccgtgtat gccaacaatg 780
gctctgtgct tcagggcacc tctgtggctt ctgtgtacca tgggaaaatt ctcataggca 840
ccgtatttca caaaactctg tactgtgagc tctagactct agatagtaaa aaaaaaaaaa 900

aaaaagtctt t 911

<210> 45
<211> 1170
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511726CB1

<400> 45
gggggcggct ccgcggggcc gggcgaggag tccagagagc catggccgcc ctgcgtgtcc 60
tgctgtcctg cgcccgcggc ccgctgaggc cccggttcg ctgtcccgcc tggcgtccct 120
tcgcctcggg tgctaacttt gagtacatca tcgcagaaaa aagagggaag aataacaccg 180
tggggttgat ccaactgaac cgccccaagg ccctcaatgc actttgcgat ggcctgattg 240
acgagctcaa ccaggccctg aagatcttcg aggaggaccc ggccgtgggg gccattgtcc 300
tcaccggcgg ggataaggcc tttgcagctg gagctgatat caaggaaatg cagaacctga 360
gtttccagga ctgttactcc agcaagttct tgaagcactg ggaccacctc acccaggtca 420
agaagccagt catcgctgct gtcaatggct atgcctttgg cgggggctgt gagcttgcca 480
tgatgtgtga tatcatctat gccggtgaga aggcccagtt tgcacagccg gagatcttaa 540
taggaaccat cccaggtgcg ggcggcaccc agagactcac ccgtgctgtt gggaagtcgc 600
tggcgatgga gatggtcctc accggtgacc ggatctcagc ccaggacgcc aagcaagcag 660
cttttgaaat gacattaaca gaaggaagta agttggagaa gaaactcttt tattcaacct 720
ttgccactga tgaccggaaa gaagggatga ccgcgtttgt ggaaaagaga aaggccaact 780
tcaaagacca gtgagaacca gctgcccctg cttcacacct ctgcttggag aggacaagtg 840
cagcctgtca gttttagaag caagtaaatc atcctctttt caagagcagt gtccgtggtg 900
tgcagttcct ctccaattgc tgcgtggtcg tggcccgacc tctcacggca tgacagcctt 960
cgtcacccag cctgtgaggg tcctgactgg agcaccttct aaatctaaga ttctgctgag 1020
gagcccccgc tggtccctct gggcatgctg tgctcggacg gaaagcgggg cctgtgggtc 1080
cttgtgtccc tgccgctgaa gaatggggct gctctgaggg aaacgctgtc tgctgccttc 1140
atacagatgc tgattaaagt gatagcgatt 1170

<210> 46
<211> 976
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511057CB1

<400> 46
accaccgtgg cctgcgacga aatggcgaaa agtcttttga agacagcctc tctgtctgga 60
aggacaaaat tgctacatca aacaggattg tcactttata gtacatccca tggattttat 120
gaggaagaag tgaaaaaaac acttcagcag tttcctggtg gatccattga ccttcagaag 180
gaagacaatg gcattggcat tcttactctg aacaatccaa gtagaatgaa tgccttttca 240
ggtgttatga tgctacaact tctggaaaaa gtaattgaat tggaaaattg gacagagggg 300
aaaggcctca ttgtccgtgg ggcaaaaaat actttctctt caggatctga tctgaatgct 360
gtgaaatcac taggaactcc agagacttcc tttaataagt gttgcgctgg ttcaaggttg 420
ggcattgggt ggaggagcag aatttactac agcatgtgat ttcaggttaa tgactccaga 480
gagtaagatc agattcgtcc acaaagagat gggcataata ccaagctggg gtggcaccac 540
ccggctagtt gaaataatcg gaagtagaca agctctcaaa gtgttgagtg ggcccttaa 600
actggattca aaaaatgctc taaacatagg aatggttgaa gaggtcttgc agtcttcaga 660
tgaaactaaa tctctagaag aggcacaaga atggctaaag caattcatcc aagggccacc 720
ggaagtaatt agagctttga aaaaatctgt ttgttcaggc agagagctat atttggagga 780

agcattacag aacgaaagag atcttttagg aacagttggg gtgggcctgc aaatttagag 840
gctattgcta agaaaggaaa atttaataaa taattggttt ttcgtgtgga tgtactccaa 900
gtaaagctcc agtgactaat atgtataaat gttaaatgat attaaatatg aacatcagaa 960
ttaaaaaaaa aaaaaa 976

<210> 47
<211> 1217
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511078CB1

<400> 47
gcctgatttg agatggggtc ccaggtctcg gtggaatcgg gagctctgca cgtggtgatt 60
gtgggtgggg gctttggcgg gatcgcagca gccagccagc tgcaggccct gaacgtcccc 120
ttcatgctgg tggacatgaa ggactccttc caccacaatg tggctgctct ccgagcctcc 180
gtggagacag ggttcgccaa aaagacattc atttcttact cggtgacttt caaggacaac 240
ttccggcagg ggctagtagt ggggatagac ctgaagaacc agatggtgct gctgcagggt 300
ggcgaggtcc agcgctcacg gttcatcgtg gtggtgggag gaggctcggc tggagtggag 360
atggcagcag agattaaaac agaatatcct gagaaagagg tcactctcat tcactcccaa 420
gtggccctgg ctgacaagga gctcctgccc tccgtccggc aggaagtgaa ggagatcctc 480
ctccggaagg gcgtgcagct gctgctgagt gagcgggtga gcaatctgga ggagctgcct 540
ctcaatgagt atcgagagta catcaaagtg cagacggaca aaggcacaga ggtggccacc 600
aacctggtga ttctctgcac cggcatcaag atcaacagct ccgcctaccg caaagcattt 660
gagagcagac tagccagcag tggtgctctg agagtgaacg agcacctcca ggtggagggc 720
cacagcaacg tctacgccat tggtgactgt gccgacgtga ggacgcccaa gatggcctat 780
cttgccggcc tccacgccaa catcgccgtg gccaacatcg tcaactctgt gaagcagcgg 840
cctctccagg cctacaagcc gggtgcactg acgttcctcc tgtccatggg gagaaatgac 900
ggtgtgggcc aaatcagtgg cttctatgtg ggccggctca tggttcggct gaccaagagc 960
cgggacctgt tcgtctctac gagctggaaa accatgaggc agtctccacc ttgatggaga 1020
ggccaggcgg gagaactacc gcagcaggtg ggcgtacgga ctgcttggcg catggcaccc 1080
gcctggcaag tgctagaact aatgctattc ttctggaata agatgccaat gatgtggtgg 1140
ctagaaatgc cacttgtata aaacaaaaat gggagagaga gaggtattaa acaaataccc 1200
ccccttttaaa aaaaaaa 1217

<210> 48
<211> 2520
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511505CB1

<400> 48
cttcgacccc tggctagagg gtaggcggcg tggagcagcg cgcgcaagcg aggccagggg 60
aaggtgggcg caggtgaggg gccgaggtgt gcgcaggact ttagccggtt gagaaggatc 120
aagcaggcat ttggagcaca ggtgtctaga aacttttaag gggccggttc aagaaggaaa 180
agttcccttc tgctgtgaaa ctatttggca agaggctgga gggcccaatg gctgcaaaat 240
cgcaacccaa cattcccaaa gccaagagtc tagatggcgt caccaatgac agaaccgcat 300
ctcaagggca gtggggccgt gcctgctgcg ccctgaccgg ccctgtggtg gacatcgtca 360
ccggacatgc tcggctctcg gacatctggg ccaagactcc acctataacg aggaaagccg 420
cccagctcta taccttgtgg gtcaccttcc aggtgcttct gtacacgtct ctccctgact 480
tctgccataa gtttctaccc ggctacgtag gaggcatcca ggagggggcc gtgactcctg 540

caggggttgt gaacaagtat cagatcaacg gcctgcaagc ctggctcctc acgcacctgc 600
tctggtttgc aaacgctcat ctcctgtcct ggttctcgcc caccatcatc ttcgacaact 660
ggatcccact gctgtggtgc gccaacatcc ttggctatgc cgtctccacc ttcgccatgg 720
tcaagggcta cttcttcccc accagcgcca gagactgcaa attcacaggc aatttctttt 780
acaactacat gatgggcatc gagtttaacc ctcggatcgg gaagtggttt gacttcaagc 840
tgttcttcaa tgggcgcccc gggatcgtcg cctggaccct catcaacctg tccttcgcag 900
cgaagcagcg ggagctccac agccatgtga ccaatgccat ggtcctggtc aacgtcctgc 960
aggccatcta cgtgattgac ttcttctgga acgaaacctg gtacctgaag accattgaca 1020
tctgccatga ccacttcggg tggtacctgg gctggggcga ctgtgtctgg ctgccttatc 1080
tttacacgct gcagggtctg tacttggtgt accacccegt gcagctgtcc accccgcacg 1140
ccgtgggcgt cctgctgctg ggcctggtgg gctactacat cttccgggtg gccaaccacc 1200
agaaggacct gttccgccgc acggatgggc gctgcctcat ctggggcagg aagcccaagg 1260
tcatcgagtg ctcctacaca tccgccgacg ggcagaggca ccacagcaag ctgctggtgt 1320
cgggcttctg ggcgtggccc cgccacttca actacgtcgg cgacctgatg ggcagcctgg 1380
cctactgcct ggcctgtggc ggtggccacc tgctgcccta cttctacatc atctacatgg 1440
ccatcctgct gacccaccgc tgcctccggg acgagcaccg ctgcgccagc aagtacggcc 1500
gggactggga gcgctacacc gccgcagtgc cttaccgcct gctgcctgga atcttctaag 1560
ggcacgccct agggagaagc cctgtggggc tgtcaagagc gtgttctgcc aggtccatgg 1620
gggctggcat cccagctcca actcgaggag cctcagtttc ctcatctgta aactggagag 1680
agcccagcac ttggcaggtg tccagtacct aatcacgctc tgttccttgc ttttgccttc 1740
aagggaattc cgagtgtcca gcactgccgt attgccagca cagacggatt ttctctaatc 1800
agtgtccctg ggcaggaggg atgacccagt cacctttact agtcctttgg agacaattta 1860
cctgtattag gagcccaggc cacgctacac tctgcccaca ctggtgagca ggaggtcttc 1920
ccacgccctg tcattaggct gcatttactc ttgctaaata aaagtgggag tggggcgtgc 1980
gcgttatcca tgtattgcct ttcagctcta gatcccctc ccctgcctgc tctgcagtcg 2040
tgggtggggc ccgtgcgccg tttctccttg gtagcgtgca cggtgttgaa ctgggacact 2100
ggggagaaag gggctttcat gtcgtttcct tcctgctcct gctgcacagc tgccaggagt 2160
gctctgcctg gagtctgcag acctcagaga ggtcccagca ctggctgtgg cctttcaggt 2220
gtaggcaggt gggctctgct tcccgattcc ctgtgagcgc ccaccctctc gaaagaattt 2280
tctgcttgcc ctgtgactgt gcagactctg gctcgagcaa cccggggaac ttcaccctca 2340
ggggcctccc acaccttctc cagcgaggag gtctcagtcc cagcctcggg agggcacctc 2400
cttttctgtg ctttcttccc tgaggcattc ttcctcatcc ctagggtgtt gtgtagaact 2460
ctttttaaac tctatgctcc gagtagagtt catctttata ttaaacttcc cctgttcaaa 2520

<210> 49
<211> 1783
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511552CB1

<400> 49
gggaagagag atctgcagac tcccagggaa cctacagctt tctccttcac agcattgcag 60
gagatggtgt ctcgctcttt tgcccaggct ggactgcagt ggcacaacct cagctcattg 120
cggcctccgc ctcctgggct caagcgattc tcctgcgtct gcctcccaag tagttgggac 180
tacaggagcc gaatcagaaa tgtcatcctc aggcacgcca gacttacctg tcctactcac 240
cgatttgaag attcaatata ctaagatctt cataaacaat gaatggcatg attcagtgag 300
tggcaagaaa tttcctgtct ttaatcctgc aactgaggag gagctctgcc aggtagaaga 360
aggagataag gaggatgttg acaaggcagt gaaggccgca agacaggctt ttcagattgg 420
atccccgtgg cgtactatgg atgcttccga gagggggcga ctattataca agttggctga 480
tttaatcgaa agagatcgtc tgctgctggc gacaatggag tcaatgaatg gtggaaaact 540
ctattccaat gcatatctga atgatttagc aggctgcatc aaaacattgc gctactgtgc 600
aggttgggct gacaagatcc agggccgtac aataccaatt gatggaaatt tttttacata 660
tacaagacat gaacctattg gtgtatgtgg ccaaatcatt ccttggaatt tcccgttggt 720

tatgctcatt tggaagatag ggcctgcact gagctgtgga aacacagtgg ttgtcaaacc 780
agcagagcaa actcctctca ctgctctcca cgtggcatct ttaataaaag aggcccgtgg 840
gggaataaag gctactttgt ccagcccaca gtgttctcta atgttacaga tgagatgcgc 900
attgccaaag aggagatttt tggaccagtg cagcaaatca tgaagtttaa atctttagat 960
gacgtgatca aaagagcaaa caatactttc tatggcttat cagcaggagt gtttaccaaa 1020
gacattgata aagccataac aatctcctct gctctgcagg caggaacagt gtgggtgaat 1080
tgctatggcg tggtaagtgc ccagtgcccc tttggtggat tcaagatgtc tggaaatgga 1140
agagaactgg gagagtacgg tttccatgaa tatacagagg tcaaaacagt cacagtgaaa 1200
atctctcaga agaactcata aagaaaatac aagagtggag agaagctctt caatagctaa 1260
gcatctcctt acagtcacta atatagtaga ttttaaagac aaaatttttc ttttcttgat 1320
ttttttaaac ataagctaaa tcatattagt attaatacta cccatagaaa acttgacatg 1380
tagcttcttc tgaaagaatt atttgccttc tgaaatgtga cccccaagtc ctatcctaaa 1440
taaaaaaaga caaattcgga tgtatgatct ctctagcttt gtcatagtta tgtgattttc 1500
ctttgtagct acttttgcag gataataatt ttatagaaaa ggaacagttg catttagctt 1560
ctttccctta gtgactcttg aagtacttaa catacacgtt aactgcagag taaattgctc 1620
tgttcccagt agttataaag tccttggact gttttgaaaa gtttcctagg atgtcatgtc 1680
tgcttgtcaa aagaaataat ccctgtaata tttagctgta aactgaatat aaagcttaat 1740
aaaaacaacc ttgcatgaaa ataaaaaaaa aaaaactcag aaa 1783

<210> 50
<211> 870
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511722CB1

<400> 50
ggcatggcgc cgctgcgctt ctccgccaat ctgtcctggc tattccccga tctctccggc 60
ctcccccgcgc gggtgcgggc cgcgggcagc tcgggcttcg aggccgtcga ggtggcctgg 120
ccgtacgcgg agacgcctga ggcgctggcg cgcgccgcgc gagaagcggg gctgcggctt 180
gtactgatca acacgccccc gggagaccaa gagaaggggg aaatggggct gggggccgtc 240
cccgggagac aggcggcctt ccgagaggga ctggagcagg ccgtgcggta tgccaaagcc 300
ctgggctgtc ccaggatcca cctgatggct ggccgagtac cccagggagc tgatcgaata 360
gcagtcaagg ctgagatgga ggccgttttt ctggagaacc tgaggcatgc agctggggtt 420
ttggctcagg aggacctcgt gggactgctg gagcccatca acacccgcat cactgacccc 480
cagtacttcc tggacacgcc ccagcaggcg gcagccatct tacagaagga catattccac 540
tggcagatca tggatgggaa cctgacagga aacatccggg agttcctgcc cattgttggg 600
catgtgcagg tggcacaggt cccaggccga ggggagccca gcagccccgg agagctgaat 660
ttcccctatc tgtttcaact gctggaagat gaaggctaca aaggcttcgt gggctgagac 720
acagtagagg gcttgagttg gctacgttca tactgggata ggcggggcca cccagaggct 780
ggccagtgag ggcccgcaca ccacccacgt gcctccagac agcgagtgac atcccatctc 840
ctcctctgca ttaaagatga cctgctgaaa 870

<210> 51
<211> 1572
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511489CB1

<400> 51
ggggcgtcgc cgtactaggc ctgccccctg tccggccagc ccctcgaagc acctactcca 60

caggtccagc cggccggtga gcgcctgggg accgcagagg tgagagtcgc gcccgggagt 120
ccgccgcctg cgccaggatg gagttcgtga aatgccttgg ccaccccgaa gagttctaca 180
acctggtgcg cttccggatc gggggcaagc ggaaggtgat gcccaagatg gaccaggact 240
cgctcagcag cagcctgaaa acttgctaca agtatctcaa tcagaccagt cgcagtttcg 300
cagctgttat ccaggcgctg gatggggaaa tgcgcaacgc agtgtgcata ttttatctgg 360
ttctccgagc tctggacaca ctggaagatg acatgaccat cagtgtggaa aagaaggtcc 420
cgctgttaca caactttcac tctttccttt accaaccaga ctggcggttc atggagagca 480
aggagaagga tcgccaggtg ctggaggact tcccaacgat ctcccttgag tttagaaatc 540
tggctgagaa ataccaaaca gtgattgccg acatttgccg gagaatgggc attgggatgg 600
cagagttttt ggataagcat gtgacctctg aacaggagtg ggacaagtac tgccactatg 660
ttgctgggct ggtcggaatt ggcctttccc gtcttttctc agcctcagag tttgaagacc 720
ccttagttgg tgaagataca gaacgtgcca actctatggg cctgtttctg cagaaaacaa 780
acatcatccg tgactatctg gaagaccagc aaggaggaag agagttctgg cctcaagagg 840
tgatggccat tgccactttg gctgcctgtt ataataacca gcaggtgttc aaaggggcag 900
tgaagattcg gaaagggcaa gcagtgaccc tgatgatgga tgccaccaat atgccagctg 960
tcaaagccat catatatcag tatatggaag agatttatca tagaatcccc gactcagacc 1020
catcttctag caaaacaagg cagatcatct ccaccatccg gacgcagaat cttcccaact 1080
gtcagctgat ttcccgaagc cactactccc ccatctacct gtcgtttgtc atgcttttgg 1140
ctgccctgag ctggcagtac ctgaccactc tctcccaggt aacagaagac tatgttcaga 1200
ctggagaaca ctgatcccaa atttgtccat agctgaagtc caccataaag tggatttact 1260
ttttttcttt aaggatggat gttgtgttct ctttattttt ttcctactac tttaatccct 1320
aaaagaacgc tgtgtggctg ggacctttag gaaagtgaaa tgcaggtgag aagaacctaa 1380
acatgaaagg aaagggtgcc tcatcccagc aacctgtcct tgtgggtgat gatcactgtg 1440
ctgcttgtgg ctcatggcag agcattcagt gccacggttt aggtgaagtc gctgcatatg 1500
tgactgtcat gagatcctac ttagtatgat cctggctaga atgataatta aaagtattta 1560
atttgaaaaa aa 1572

<210> 52
<211> 1042
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511497CB1

<400> 52
cggcaggcgc cagtcgcagg tgtgctgctg aggcgtgaga atggcgtccc gcggccggcg 60
tccggagcat ggcggacccc cagagctgtt ttatgacgag acagaagccc ggaaatacgt 120
tcgcaactca cggatgattg atatccagac caggatggct gggcgagcat tggagcttct 180
ttatctgcca gagaataagc cctgttacct gctggatatt ggctgtggca ctgggctgag 240
tggaagttat ctgtcagatg aagggcacta ttgggtgggc ctggatatca gccctgccat 300
gctggcattt ctgctgtgca gtggctctgt aatgctaaca agaagtctga aaaccctgcc 360
aagcgcctgt actgcttttt tgcttctctt ttttctgttc tcgtccgggg atcccgagct 420
gtcctgcagc tgtaccctga gaactcagag cagttggagc tgatcacaac ccaggccaca 480
aaggcaggct tctccggtgg catggtggta gactacccta acagtgccaa agcaaagaaa 540
ttctacctct gcttgttttc tgggccttcg acctttatac cagaggggct gagtgaaaat 600
caggatgaag ttgaacccag ggagtctgtg ttcaccaatg agaggttccc attaaggatg 660
tcgaggcggg gaatggtgag gaagagtcgg gcatgggtgc tggagaagaa ggagcggcac 720
aggcgccagg gcagggaagt cagacctgac acccagtaca ccggccgcaa gcgcaagccc 780
cgcttctaag tcaccacgcg gttctggaaa ggcacttgcc tctgcacttt tctatattgt 840
tcagctgaca aagtagtatt ttagaaaagt tctaaagtta taaaaatgtt ttctgcagta 900
aaaaaaaagt tctctgggcc gggcgtggtg gctcacacct gtaatcccag caccttggga 960
ggctgaggtg ggaggatcat ttgaggccag gagtttgaga cctgcctggg caacataatg 1020
aaacttcctt tccagggaga aa 1042

<210> 53
<211> 921
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511498CB1

<400> 53
cggcaggcgc cagtcgcagg tgtgctgctg aggcgtgaga atggcgtccc gcggccggcg 60
tccggagcat ggcggacccc cagagctgtt ttatgacgag acagaagccc ggaaatacgt 120
tcgcaactca cggatgattg atatccagac caggatggct gggcgagcat tggagcttct 180
ttatctgcca gagaataagc cctgttacct gctggatatt ggctgtggca ctgggctgag 240
tggaagttat ctgtcagatg aagggcacta ttgggtgggc ctggatatca gccctgccat 300
gctggatgag gctgtggacc gagagataga gggagacctg ctgctggggg atatgggcca 360
gggcatccca ttcaagccag gcacatttga tggttgcatc agcatttctg ctgtgcagtg 420
gctctgtaat gctaacaaga agtctgaaaa ccctgccaag cgcctgtact gctttttttgc 480
ttctcttttt tctgttctcg tccggggatc ccgagctgtc ctgcagctgt accctgagaa 540
ctcagagcag ttggagctga tcacaaccca ggccacaaag gcaggcttct ccggtggcat 600
ggtggtagac taccctaaca gtgccaaagc aaagaaattc tacctctgct tgttttctgg 660
gccttcgacc tttataccag aggttcccat taaggatgtc gaggcgggga atggtgagga 720
agagtcgggc atgggtgctg gagaagaagg agcggcacag gcgccagggc agggaagtca 780
gacctgacac ccagtacacc ggccgcaagc gcaagccccg cttctaagtc accacgcggt 840
tctggaaagg cacttgcctc tgcacttttc tatattgttc agctgacaaa gtagtatttt 900
agaaaagttc taaagttata a 921

<210> 54
<211> 1754
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511612CB1

<400> 54
cttggaggct gagatgtggt tctgcgcgtg tgcggacggc tgtctgttaa ctccgcggtc 60
agttcccgga ctggtggctg gtctgcaggg ttgacctgcg caatgcagag gctgcaggta 120
gtgctgggcc acctgagggg tccggccgat tccggctgga tgccgcaggc cgcgccttgc 180
ctgagcggtg ccccgcaggc ctcggccgcg gacgtggtgg tggtgcacgg gcggcgcacg 240
gccatctgcc gggcgggccg cggcggcttc aaggacacca cccccgacga gcttctctcg 300
gcagtcatga ccgcggttct caaggacgtg aatctgaggc cggaacagct gggggacatc 360
tgtgtcggaa atgtgctgca gcctggggcc ggggcaatca tggcccgaat cgcccagttt 420
ctgagtgaca tcccggagac tgtgcctttg tccactgtca atagacagtg ttcgtcgggg 480
ctacaggcag tggccagcat agcagggtgg agtccatgtc cctggctgac agagggaacc 540
ctggaaatat tacttcgcgc ttgatggaga aggagaaggc cagagattgc ctgattccta 600
tggggataac ctctgagaat gtggctgagc ggtttggcat ttcacgggag aagcaggata 660
cctttgccct ggcttcccag cagaaggcag caagagccca gagcaagggc tgtttccaag 720
ctgagattgt gcctgtgacc accacggtcc atgatgacaa gggcaccaag aggagcatca 780
ctgtgaccca ggatgagggt atccgcccca gcaccaccat ggagggcctg gccaaactga 840
agcctgcctt caagaaagat ggttctacca cagctggaaa ctctagccag gtgagtgatg 900
gggcagctgc catcctgctg gcccggaggt ccaaggcaga agagttgggc cttcccatcc 960
ttggggtcct gaggtcttat gcagtggttg gggtcccacc tgacatcatg ggcattggac 1020
ctgcctatgc catcccagta gctttgcaaa aagcagggct gacagtgagt gacgtggaca 1080
tcttcgagat caatgaggcc tttgcaagcc aggctgccta ctgtgtggag aagctacgac 1140

tccccccctga gaaggtgaac cccctggggg gtgcagtggc cttagggcac ccactgggct 1200
gcactggggc acgacaggtc atcacgctgc tcaatgagct gaagcgccgt gggaagaggg 1260
catacggagt ggtgtccatg tgcatcggga ctggaatggg agccgctgcc gtctttgaat 1320
accctgggaa ctgagtgagg tcccaggctg gaggcgctac gcagacagtc ctgctgctct 1380
agcagcaagg cagtaacacc acaaaagcaa aaccacatgg gaaaactcag cactggtggt 1440
ggtggcagtg gacagatcaa ggcacttcaa ctcatttgga aaatgtgaac actgatgaca 1500
tggtatagga gtgggtgggg tgttgagcca cccatcagac cctctttagc tgtgcaagat 1560
aaaagcagcc tgggtcaccc aggccacaag gccatggtta attcttaagg caaggcaaat 1620
ccatggatga gaagtgcaat gggcatagta aaagtgcatg aatttaaaaa aaaaaaaaaa 1680
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaagaa aaaaaaaaaa aaaaaaggag 1740
gaccgggccg ttcg 1754

<210> 55
<211> 2262
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511624CB1

<400> 55
cttcagcagg agccggttcc ctgtgtgtgt gtccgctcgc cctctgctcc gtcctgcggc 60
tgcccactgc cctcctacgg tccaccatgg ccctgctgca ctccggccgc gtcctccccg 120
ggatcgccgc cgccttccac ccgggcctcg ccgccgcggc ctctgccaga gccagctcct 180
ggtggaccca tgtggaaatg ggacctccag atcccattct gggagtcact gaagccttta 240
agagggacac caatagcaaa aagatgaatc tgggagttgg tgcctaccgg gatgataacg 300
gaaagcctta cgtgctgcct agcgtccgca aggcagaggc ccagattgcc gcaaaaaatt 360
tggacaagga atacctgccc attgggggac tggctgaatt ttgcaaggca tctgcagaac 420
tagccctggg tgagaacagc gaagtcttga agagtggccg gtttgtcact gtgcagacca 480
tttctggaac tggagcctta aggatcggag ccagttttct gaaaatacca gagcagagtg 540
ttcttcttct gcatgcctgc gcccacaatc ccacgggagt ggacccgcgt ccggaacagt 600
ggaaggaaat agcaacagtg gtgaagaaaa ggaatctctt tgcgttcttt gacatggcct 660
accaaggctt tgccagtggt gatggtgata aggatgcctg ggctgtgcgc cacttcatcg 720
aacagggcat taatgtttgt ctctgccaat catatgccaa gaacatgggc ttatatggtg 780
agcgtgtagg agccttcact atggtctgca aagatgcgga tgaagccaaa agggtagagt 840
cacagttgaa gatcttgatc cgtcccatgt attccaaccc tcccctcaat ggggcccgga 900
ttgctgctgc cattctgaac accccagatt tgcgaaaaca atggctgcaa gaagtgaaag 960
gcatggctga ccgcatcatt ggcatgcgga ctcaactggt ctccaacctc aagaaggagg 1020
gttccaccca caattggcaa cacatcaccg accaaattgg catgttctgt ttcacagggc 1080
taaagcctga acaggtggag cggctgatca aggagttctc catctacatg acaaaagatg 1140
gccgcatctc tgtggcaggg gtcacctcca gcaacgtggg ctaccttgcc catgccattc 1200
accaggtcac caagtaatgt ccctggtgcg aggaaacaga gacaaccttt ctgtcttcag 1260
cctctgctat tgagagcttc acacagacaa tgagagaggg tggatggtgg tgagtggatc 1320
atttctttca gccacagtgt gtaacactca gcatttgaat gtttctcaga aaagaacatg 1380
tagtgacaca gggcagaggc atccatggct ggcgtctgga atattaaacc aaactctccc 1440
cggtcctttt ttctccaact tttctcaaag agtttacatg tgcaagaaag tcatcgcacc 1500
aaaaaacctg tcaattatgc cattgcaata tttcagaagc tttaactgaa gtgtcaggtt 1560
cctcgtgaga aacagcacac gttagaggct ttgagagaag gcctagttct gtcatgagta 1620
gtcggcctcg tgtctgtcct cccatcttgg aacaacctta tcaacaggcc gcactgcaga 1680
aatgatgttt tatgaaaacc aatgaggctg ctgcctcact ccagcaaggg aaataatgca 1740
gtttcctgtc ttatttaaga aaaagagaag gctctctttt ctcccttgtc attgccgttc 1800
ttttccttac acgcaaagat tttttaacta ttgcagattt tcatcccatt ctactgcttg 1860
attgaccatc aactccatcc tatcgagatt tatttaagaa tgaagaacat aattttctgc 1920
tgatgccgta ccctcaccct tttcagcaaa gaatagtgga gagtaggaaa ctgtacttta 1980
tctcggcatc ctcttgaatg atagtgcaag tttctccagt tgggatgttg tctctgcccg 2040

gttggacctc ctccctttgt tgaatgtggt gtgcagcctc tcatctcaca ctgtgagtcc 2100
agcggcgcag ggtggtacca ggaaagagga tattctaggc tttgcgtgct gctagctggg 2160
ttcaggcttc acccactgga aagaaccacc atctgctcta accatgtaga cttattgcgg 2220
cctggtttct ctgttacaat aaaattactg tagacccaaa aa 2262

<210> 56
<211> 2295
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511626CB1

<400> 56
cttcagcagg agccggttcc ctgtgtgtgt gtccgctcgc cctctgctcc gtcctgcggc 60
tgcccactgc cctcctacgg tccaccatgg ccctgctgca ctccggccgc gtcctccccg 120
ggatcgccgc cgccttccac ccgggcctcg ccgccgcggc ctctgccaga gccagctcct 180
ggtggaccca tgtggaaatg ggacctccag atcccattct gggagtcact gaagccttta 240
agagggacac caatagcaaa aagatgaatc tgggagttgg tgcctaccgg gatgataacg 300
gaaagcctta cgtgctgcct agcgtccgca agtttgtcac tgtgcagacc atttctggaa 360
ctggagcctt aaggatcgga gccagttttc tgcaaagatt ttttaagttc agccgagatg 420
tctttctgcc caaaccaacc tggggaaacc acacacccat cttcagggat gctggcatgc 480
agctacaagg ttatcggtat tatgacccca agacttgcgg ttttgacttc acaggcgctg 540
tggaggatat ttcaaaaata ccagagcaga gtgttcttct tctgcatgcc tgcgcccaca 600
atcccacggg agtggacccg cgtccggaac agtggaagga aatagcaaca gtggtgaaga 660
aaaggaatct ctttgcgttc tttgacatgg cctaccaagg ctttgccagt ggtgatggtg 720
ataaggatgc ctgggctgtg cgccacttca tcgaacaggg cattaatgtt tgtctctgcc 780
aatcatatgc caagaacatg ggcttatatg gtgagcgtgt aggagccttc actatggtct 840
gcaaagatgc ggatgaagcc aaaagggtag agtcacagtt gaagatcttg atccgtccca 900
tgtattccaa ccctcccctc aatggggccc ggattgctgc tgccattctg aacaccccag 960
atttgcgaaa acaatggctg caagaagtga aaggcatggc tgaccgcatc attggcatgc 1020
ggactcaact ggtctccaac ctcaagaagg agggttccac ccacaattgg caacacatca 1080
ccgaccaaat tggcatgttc tgtttcacag ggctaaagcc tgaacaggtg gagcggctga 1140
tcaaggagtt ctccatctac atgacaaaag atggccgcat ctctgtggca ggggtcacct 1200
ccagcaacgt gggctacctt gcccatgcca ttcaccaggt caccaagtaa tgtccctggt 1260
gcgaggaaac agagacaacc tttctgtctt cagcctctgc tattgagagc ttcacacaga 1320
caatgagaga gggtggatgg tggtgagtgg atcatttctt tcagccacag tgtgtaacac 1380
tcagcatttg aatgtttctc agaaaagaac atgtagtgac acagggcaga ggcatccatg 1440
gctggcgtct ggaatattaa accaaactct ccccggtcct tttttctcca acttttctca 1500
aagagtttac atgtgcaaga aagtcatcgc accaaaaaac ctgtcaatta tgccattgca 1560
atatttcaga agctttaact gaagtgtcag gttcctcgtg agaaacagca cacgttagag 1620
gctttgagag aaggcctagt tctgtcatga gtagtcggcc tcgtgtctgt cctcccatct 1680
tggaacaacc ttatcaacag gccgcactgc agaaatgatg ttttatgaaa accaatgagg 1740
ctgctgcctc actccagcaa gggaaataat gcagtttcct gtcttattta agaaaaagag 1800
aaggctctct tttctccctt gtcattgccg ttcttttcct tacacgcaaa gattttttaa 1860
ctattgcaga ttttcatccc attctactgc ttgattgacc atcaactcca tcctatcgag 1920
atttatttaa gaatgaagaa cataattttc tgctgatgcc gtaccctcac ccttttcagc 1980
aaagaatagt ggagagtagg aaactgtact ttatctcggc atcctcttga atgatagtgc 2040
aagtttctcc agttgggatg ttgtctctgc ccggttggac ctcctccctt tgttgaatgt 2100
ggtgtgcagc ctctcatctc acactgtgag tccagcggcg cagggtggta ccaggaaaga 2160
ggatattcta ggctttgcgt gctgctagct gggttcaggc ttcacccact ggaaagaacc 2220
accatctgct ctaaccatgt agacttattg cggcctggtt tctctgttac aataaaatta 2280
ctgtagaccc aaaaa 2295

<210> 57

<211> 1108
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512885CB1

<400> 57
gttgattgac tatggtctct ccggctacca ggaagagtct gccgaaggtg aaggccatgg 60
acttcatcac ctccacagcc atcctgcccc tgctgttcgg ctgcctgggc gtcttcggcc 120
tcttccggct gctgcagtgg gtgcgcggga aggcctacct gcggaatgct gtggtggtga 180
tcacaggcgc cacctcaggg ctgggcaaag aatgtgcaaa agtcttctat gctgcgggtg 240
ctaaactggt gctctgtggc cggaatggtg gggccctaga agagctcatc agagaactca 300
ccgcttctca tgccaccaag gtgcagacac acaagcctta cttggtgacc ttcgacctca 360
cagactctgg ggccatagtt gcagcagcag ctgagatcct gcagtgcttt ggctatgtcg 420
acatacttgt caacaatgct gggatcagct accgtggtac catcatggac accacagtgg 480
atgtggacaa gagggtcatg gagacaaact actttggccc agttgctcta acgaaagcac 540
tcctgccctc catgatcaag aggaggcaag gccacattgt cgccatcagc agcatccagg 600
gcaagatgag cattcctttt cgatcagcat ttatggacac caccacagcc cagggccgaa 660
gccctgtgga ggtggcccag gatgttcttg ctgctgtggg gaagaagaag aaagatgtga 720
tcctggctga cttactgcct tccttggctg tttatcttcg aactctggct cctgggctct 780
tcttcagcct catggcctcc agggccagaa aagagcggaa atccaagaac tcctagtact 840
ctgaccagcc agggccaggg cagagaagca gcactcttag gcttgcttac tctacaaggg 900
acagttgcat ttgttgagac tttaatggag atttgtctca caagtgggaa agactgaaga 960
aacacatctc gtgcagatct gctggcagag gacaatcaaa aacgacaaca agcttcttcc 1020
cagggtgagg ggaaacactt aaggaataaa tatggagctg gggtttaaca ctaaaaacta 1080
gaaataaaca tctcaaacag taaaaaaa 1108

<210> 58
<211> 1170
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511965CB1

<400> 58
ggaaggccga ggtgggcgga tcacctgagc tcaggagttt gagaccagcc tgtctctact 60
aacaatataa aaattagctg ggagtcacgg tgggcgcctg taatcccagc tactcgggag 120
gctgaggcag gagaattgct tgaacccagg agacagaggt tgtagtgagc tgagatcgca 180
ccactgcact ctagccttgg caacagtgca agactgtctc aaaaacagca acagagagca 240
ggacgtgaga cttctacctg ctcactcaga atcatttctg caccaaccat ggccacgttt 300
gtggagctca gtacaaaagc caagatgccc attgtgggcc tgggcacttg gaggtctctt 360
ctcggcaaag tgaaagaagc ggtgaaggtg gccattgatg caggatatcg gcacattgac 420
tgtgcctatg tctatcagaa tgaacatgaa gtgggggaag ccatccaaga gaagatccaa 480
gagaaggctg tgatgcggga ggacctgttc atcgtcagca aggtgtggcc cactttcttt 540
gagagacccc ttgtgaggaa agcctttgag aagaccctca aggacctgaa gctgagctat 600
ctggacgtct atcttattca ctggccacag ggattcaaga ctggggatga ctttttcccc 660
aaagatgata aaggtaatat gatcagtgga aaaggaacgt tcttggatgc ctgggagatc 720
gagaggctct tgaacaaacc tggactgaaa tataaaccag tgactaacca ggttgagtgt 780
cacccatacc tcacgcagga gaaactgatc cagtactgcc actccaaggg catcaccgtt 840
acggcctaca gcccccctggg ctctccggat agaccttggg ccaaacctga ggacccttcc 900
ctgctggagg atcccaagat taaggagatt gctgcaaagc acaaaaaaac cacagcccag 960
gttctgatcc gtttccatat ccagaggaat gtgacagtga tccccaagtc tatgacacca 1020

gcacacattg ttgagaacat tcaggtcttt gactttaaat tgagtgatga ggagatggca 1080
accatactca gcttcaacag aaactggagg gcctttgact tcaaggaatt ctctcatttg 1140
gaggactttc ccttcgatgc agaatattga 1170

<210> 59
<211> 2506
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512403CB1

<400> 59
ctccgactgg cagtgggact cagcgggcgt ggaggtcgcg gctgagcgag cgagccctgg 60
gcgagtgaat tgtggctgtg ggttgacggt ggagacaccc cccggaggga ggcggaggga 120
agggaggcga ggcctgcacc tgcatgcttc ccgcctccca ctccccagcg cccccggacc 180
gtgcagttct ctgcaggacc aggccatgga gctcgaagtc cggcgggtcc gacaggcgtt 240
cctgtccggc cggtcgcgac ctctgcggtt tcggctgcag cagctggagg ccctgcggag 300
gatggtgcag gagcgcgaga aggatatcct gacggccatc gccgccgacc tgtgcaagag 360
tgaattcaat gtgtacagtc aggaagtcat tactgtcctt ggggaaattg attttatgct 420
tgagaatctt cctgaatggg ttactgctaa accagttaag aagaacgtgc tcaccatgct 480
ggatgaggcc tatattcagc cacagcctct gggagtggtg ctgataatcg gagcttggaa 540
ttaccccttc gttctcacca ttcagccact gataggagcc atcgctgcag gaaatgctgt 600
gattataaag ccttctgaac tgagtgaaaa tacagccaag atcttggcaa agcttctccc 660
tcagtattta gaccaggaat tttatggaga aaatataaaa gagtctcctg attatgaaag 720
gatcatcaat cttcgtcatt ttaagaggat actaagtttg cttgaaggac aaaagatagc 780
ttttggtggg gagactgatg aggccacacg ctacatagcc ccaacagtac ttaccgatgt 840
tgatcctaaa accaaggtga tgcaagaaga aatttttgga ccaattcttc caatagtgcc 900
tgtgaaaaat gtagatgagg ccataaattt cataaatgaa cgtgaaaagc ctctggctct 960
ttatgtattt tcgcataacc ataagctcat caaacggatg attgatgaga catccagtgg 1020
aggtgtcaca ggcaatgacg tcattatgca cttcacgctc aactctttcc catttggagg 1080
agtgggttcc agtgggatgg gagcttatca cggaaaacat agttttgata ctttttctca 1140
tcagcgtccc tgtttattaa aaagtttaaa gagagaaggt gctaacaaac tcagatatcc 1200
tcccaacagc cagtcaaagg tggattgggg aaaatttttt ctcttgaaac ggttcaacaa 1260
agaaaaactc ggtctcctgt tgctcacttt cctgggtatt gtagccgctg tgcttgtcaa 1320
ggcagaatat tactgaagaa tgatcctgtt caacctccta gtgcctctac tgaattattc 1380
ctcttttaaa tggttaatga accaataatt tttaaatcat accaaaaata gtaagaaaat 1440
atgcaaacac tctgtgatca aacttaaaag tcattgccat tcatcattaa taaaagttgc 1500
catttcaact acgtcccaac attccctaat agggtattca gggaacctgt cttaaattgt 1560
gcttatctaa atcttggaac tttgagctag gggaggagaa tgtattagac taaatacaaa 1620
ctgcggggtt gtaagggagt ctcagaacct cactgaatcc ttcactccag ttaatggcac 1680
tgctcacttc ctgcctctgc tgccaccatc actgtgtgaa gctttcaaga gcttggtact 1740
tcccagggct accggcagtc ctctgtagtc cagagaggtg agattagatc ttcttggttc 1800
cctgtgaggt ttcaggcact aaaactctat gtggggaagg gaggggttac tcctcctcca 1860
atgggactca aggacttgac ctccaggagt aggcccctgg tcagaagtgc catctcacca 1920
gtggtcttca ttcttcctca ttcattcttt atcatcctgt gttctgttta gttgcaacaa 1980
tctcttgtga ctaatgtcac tcaaagcatc ttgtaaatcc tagggcttcc tggaagttag 2040
ttgccaaagt catgcaagca tcacctgtca ttcttgtgtt ggagttatag aattctacat 2100
cttataaaac ctaactggca tttaaaaaat actgtggccg ggcgtggtgg ctcatgcctg 2160
taatcccagc actttgggag gccgaggtgg gaggattgct tgagtccagg aatttgagac 2220
cagcctggac aacacagtga gacctcatct ctatcaaaaa ataaaaatta gctagatgtg 2280
gtggcatgag cctgtgttcc cagctgctta ggaggctgaa gcaggaggat tgattgagcc 2340
tgcgaggcca aggctgcagc aggctgtgat tgcaccactg cacttcagct tgggcaacag 2400
agcaagaccc tgtctccgaa acaaataaaa aatactgtaa taaaagtact tataaacata 2460
ctaatcctct ttcaggaccc taaagttgca ggttagtagg tcttca 2506

<210> 60
<211> 1827
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512564CB1

<400> 60
caacgctgga gcatcccgct ctggtgccgc tgcagccggc agagatggtt gagctcatgt 60
tcccgctgtt gctcctcctt ctgcccttcc ttctgtatat ggctgcgccc caaatcagga 120
aaatgctgtc cagtggggtg tgtacatcaa ctgttcagct tcctgggaaa gtagttgtgg 180
tcacaggagc taatacaggt atcgggaagg agacagccaa agagctggct cagagaggag 240
ctcgagtata tttagcttgc cgggatgtgg aaaaggggga attggtggcc aaagagatcc 300
agaccacgac agggaaccag caggtgttgg tgcggaaact ggacctgtct gatactaagt 360
ctattcgagc ttttgctaag ggcttcttag ctgaggaaaa gcacctccac gttttgatca 420
acaatgcagg agtgatgatg tgtccgtact cgaagacagc agatggcttt gagatgcaca 480
taggagtcaa ccacttgggc tctggcgtta cgacgtattc tgtacaccct ggcacagtcc 540
aatctgaact ggttcggcac tcatctttca tgagatggat gtggtggctt ttctcctttt 600
tcatcaagac tcctcagcag ggagcccaga ccagcctgca ctgtgcctta acagaaggtc 660
ttgagattct aagtgggaat catttcagtg actgtcatgt ggcatgggtc tctgcccaag 720
ctcgtaatga gactatagca aggcggctgt gggacgtcag ttgtgacctg ctgggcctcc 780
caatagacta acaggcagtg ccagttggac ccaagagaag actgcagcag actacacagt 840
acttcttgtc aaaatgattc tccttcaagg ttttcaaaac ctttagcaca aagagagcaa 900
aaccttccag ccttgcctgc ttggtgtcca gttaaaactc agtgtactgc cagattcgtc 960
taaatgtctg tcatgtccag atttactttg cttctgttac tgccagagtt actagagata 1020
tcataatagg ataagaagac cctcatatga cctgcacagc tcattttcct tctgaaagaa 1080
actactacct aggagaatct aagctatagc agggatgatt tatgcaaatt tgaactagct 1140
tctttgttca caattcagtt cctcccaacc aaccagtctt cacttcaaga gggccacact 1200
gcaacctcag cttaacatga ataacaaaga ctggctcagg agcagggctt gcccaggcat 1260
ggtggatcac cggaggtcag tagttcaaga ccagcctggc caacatggtg aaaccccacc 1320
tctactaaaa attgtgtata tctttgtgtg tcttcctgtt tatgtgtgcc aagggagtat 1380
tttcacaaag ttcaaaacag ccacaataat cagagatgga gcaaaccagt gccatccagt 1440
ctttatgcaa atgaaatgct gcaaagggaa gcagattctg tatatgttgg taactaccca 1500
ccaagagcac atgggtagca gggaagaagt aaaaaaaaga gaaggagaat actggaagat 1560
aatgcacaaa atgaagggac tagttaagga ttaactagcc ctttaaggat taactagtta 1620
aggattaata gcaaaagaca ttaaatatgc taacatagct atggaggaat tgagggcaag 1680
cacccaggac tgatgaggtc ttaacaaaaa ccagtgtggc aaaaaaaaaa aaaaaaaaaa 1740
aaaaaaaaaa tcctaaaaac aaacaaaaaa aaaaaacaat tcttcattca gaaaaattat 1800
cttagggact gatattggta attatgg 1827

<210> 61
<211> 714
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512646CB1

<400> 61
cgtggcgccc gagtgcactg aagatggcgg ctgctgtagg acggttgctc cgagcgtcgg 60
ttgcccgaca tgtgagtgcc attccttggg gcatttctgc cactgcagcc ctcaggcctg 120
ctgcatgtgg aagaacgagc ttgacaaatt tattgtgttc tggttccagt caagcaaaat 180
tattcagcac cagttcctca tgccatgcac ctgctgtcac ccagcatgca ccctatttta 240

agggtacagc cgttgtcaat ggagagttca aagacctaag ccttgatgac tttaagggga 300
aatatttggt gcttttcttc tatcctttgg atttcacctt tgtgtgtcct acagaaattg 360
ttgcttttag tgacaaagct aacgaatttc acgatgtgaa ctgtgaagtt gtcgcagtct 420
cagtggattc ccactttagc catcttgcct ggataaatac accaagaaag aatggtggtt 480
tgggccacat gaacatcgca ctcttgtcag acttaactaa gcagatttcc cgagactacg 540
gtgtgctgtt agaaggttct ggtcttgcac taagatcaag ccaagtccag ctgcttccaa 600
agagtacttt cagaaggtaa atcagtagat cacccatgtg tatctgcacc ttctcaactg 660
agagaagaac cacagttgaa acctgctttt atcattttca agatggttat ttgt 714

<210> 62
<211> 1034
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512700CB1

<400> 62
aatggatctc agtgctgcaa gtcaccgcat acctctaagt gatggaaaca gcattcccat 60
catcggactt ggtacctact cagaacctaa atcgctatgg gctacaaatc atgtcccaga 120
gatggtccgc ccaaccctgg agaggacact cagggtcctc cagctagatt atgtggatct 180
ttacatcatt gaagtaccca tggcctttaa gccaggagat gaaatatacc ctagagatga 240
gaatggcaaa tggttatatc acaagtcaaa tctgtgtgcc acttgggagg cgatggaagc 300
ttgcaaagac gctggcttgg tgaaatccct gggagtgtcc aattttaacc gcaggcagct 360
ggagctcatc ctgaacaagc caggactcaa acacaagcca gtcagcaacc aggttgagtg 420
ccatccgtat ttcacccagc caaaactctt gaaattttgc caacaacatg acattgtcat 480
tactgcatat agccctttgg ggaccagtag gaatccaatc tgggtgaatg tttcttctcc 540
acctttgtta aaggatgcac ttctaaactc attggggaaa aggtacaata agacagcagc 600
tcaaattgtt ttgcgtttca acatccagcg aggggtggtt gtcattccta aaagctttaa 660
tcttgaaagg atcaaagaaa attttcagat ctttgacttt tctctcactg aagaagaaat 720
gaaggacatt gaagccttga ataaaaatgt ccgctttgta gaattgctca tgtggcgcga 780
tcatcctgaa tacccatttc atgatgaata ctgactgcag ggagttcctg aacagatttt 840
tcactcccac gagtgccaag acggtgcaat gggtagtccc ctagatgtga aaatgaagag 900
agagggtttt accatcctga gaagaaataa tgatggaaac atgtttaatg tttgtgcagt 960
gtaaatgact ttgactcagt cacattgaag taaaaatatt aaaatctgtt gaaataactc 1020
ttaggaaaaa aaaa 1034

<210> 63
<211> 1071
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512707CB1

<400> 63
ctgagtgcga cggcgcaggt ctgcaccatg gcggcccggc ttgtcagccg atgcggggct 60
gtgcgtgcag ctccccacag cggcccgctg gtgtcctggc gcaggtggtc cggcgcctca 120
acagacaccg tgtatgacgt ggtggtgtcg ggtggaggcc tggtgggcgc tgccatggcc 180
tgtgcccttgg gatatgatat tcactttcat gacaagaaaa tcctgttgct cgaagcaggt 240
ccaaagaaag tactggagaa attgtcagaa acttacagca acagggtcag ctccatttcc 300
cctggctctg caacgcttct cagtagtttt ggtgcctggg accatatctg caacatgaga 360
tacagagcct ttcggcgaat gcaggtgtgg gacgcctgct cagaggccct gataatgttt 420
gataaggata atttagatga catgggctat atcgtggaga atgatgtcat catgcatgct 480

ctcactaagc agttggaggc tgtgtctgac cgagtgacgg ttctctacag gagcaaagcc 540
attcgctata cctggccttg tccatttcct atggccgact ccagcccttg ggttcatatt 600
accctaggtg atggcagcac cttccagacc aaattgttga taggtgcaga tggtcacaac 660
tccggagtac ggcaggctgt tggaatccag aatgtgagct ggaactatga ccagtctgct 720
gttgtggcta ctctgcattt atcagaggcc acagaaaaca acgtagcctg gcagagattt 780
cttccctctg ggcctattgc tctgctcccg ctctcagaca ccttgagttc cttggtttgg 840
tccacgtccc atgaacatgc agcagagcta gttagcatgg atgaggaaaa atttgtggat 900
gccgttaact ctgcctttga acagattatg gcctttgcaa gcaaatgagt actcctctcc 960
taaagaaaga ttacgttgat gaaaaagaac atcctgccca ggacccatca tacatatttt 1020
caagatctta tttaatttaa taaacttact ttacattaaa aaaaaaaaaa a 1071

<210> 64
<211> 1399
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512710CB1

<400> 64
gcgcaggtct gcaccatggc ggcccggctt gtcagccgat gcggggctgt gcgtgcagct 60
ccccacagcg gcccgctggt gtcctggcgc aggtggtccg gcgcctcaac agacaccgtg 120
tatgacgtgg tggtgtcggg tggaggcctg gtgggcgctg ccatggcctg tgccttgggt 180
tttggtgcct gggaccatat ctgcaacatg agatacagag cctttcggcg aatgcaggtg 240
tgggacgcct gctcagaggc cctgataatg tttgataagg ataatttaga tgacatgggc 300
tatatcgtgg agaatgatgt catcatgcat gctctcacta agcagttgga ggctgtgtct 360
gaccgagtga cggttctcta caggagcaaa gccattcgct atacctggcc ttgtccattt 420
cctatggccg actccagccc ttgggttcat attaccctag gtgatggcag caccttccag 480
accaaattgt tgataggtgc agatggtcac aactccggag tacggcaggc tgttggaatc 540
cagaatgtga gctggaacta tgaccagtct gctgttgtgg ctactctgca tttatcagag 600
gccacagaaa acaacgtagc ctggcagaga tttcttccct ctgggcctat tgctctgctc 660
ccgctctcag acaccttgag ttccttggtt tggtccacgt cccatgaaca tgcagcagag 720
ctagttagca tggatgagga aaaatttgtg gatgccgtta actctgcctt ttggagtgat 780
gctgaccaca cggacttcat cgacacagct ggtgccatgc tgcagtatgc tgtcagcctt 840
ctgaagccca ctaaggtctc ggctcgccag ctgcccccaa gcgtagccag ggtggatgcc 900
aaaagccgag ttctgtttcc tcttgggttg ggacatgctg ctgagtacgt caggcctcgg 960
gtggcgctca ttggggatgc agcccacaga gtccatccgc ttgcaggaca gggtgtcaac 1020
atgggctttg gggatatctc cagcttggcc catcacctca gtacggcagc cttcaatggg 1080
aaggacttag gttccgtgag ccacctcaca ggttatgaaa cagaaagaca gcgtcacaac 1140
actgctcttc tggctgctac agacttacta aaaaggctct attctaccag tgcctccccg 1200
cttgtgttgc tcaggacgtg ggcttgcag gccacaaatg cagtgtctcc actcaaagaa 1260
cagattatgg cctttgcaag caaatgagta ctcctctcct aaagaaagat tacgttgatg 1320
aaaaagaaca tcctgcccag gacccatcat acatattttc aagatcttat ttaatttaat 1380
aaacttactt tacattaaa 1399

<210> 65
<211> 1634
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512884CB1

<400> 65

ccacctttct tgggtggctc tccgcctcgt cctccctccg agggccgttg gtacattcct 60
agtgactcca agcgcttaaa aggggcccgg gaggatgaac cccacagatc tgaacctgat 120
ttgtgtgtgc accgcgtctc cagcgatccc ggatccactg cgctgccagg ggcctggggg 180
tgggtctctt gctgtctctg cgacgacatc cttacgtttc ggcactctaa tgctgggttt 240
gtgcgtgtgt gtctgcttag cggtctagcg ggctgttagg ctccctcgcc cccagctcct 300
tggctcgctc agctcctcca ccgcagccca gcagtgagac gcgcgcgcag ccagctcccc 360
acgagatgga acagaccgaa gtgctgaagc cacggaccct ggctgatctg atccgcatcc 420
tgcaccagct ctttgccggc gatgaggtca atgtagagga ggtgcaggcc atcatggaag 480
cctacgagag cgaccccacc gagtgggcaa tgtacgccaa gttcgaccag tacaggtata 540
cccgaaatct tgtggatcaa ggaaatggaa aatttaatct gatgattctc tgttggggtg 600
aaggacatgg cagatgctac agggaaatct aaaggagaca ttatttgcct ggcctgacaa 660
aaaatccaat gagatggtca agaagtctga aagagtcttg agggaaaacc agtgtgccta 720
catcaatgat tccattggct tacatcgagt agagaacatc agccatacgg aacctgctgt 780
gagccttcac ttgtacagtc caccttttga tacatgccat gcctttgatc aaagaacagg 840
acataaaaac aaagtcacaa tgacattcca tagtaaattt ggaatcagaa ctccaaatgc 900
aacttcgggc tcgctggaga acaactaagg ggcaccaaac cctctgaggt tttactttaa 960
ggttcgctgt atgtttgcct tggacaaaaa ggctacctac cacgtgctat ccagtaatat 1020
acttaaataa gccaatactt agatctactg taaggcagat gctaattata aggcattaag 1080
taagcaaata gtgccctcag ctactgcaga agaaaagtcc cactgaggaa aagaaagtct 1140
tgtgattttt aaaggcaagt tttcaagtgc tctcatagtt ctatcctcta attccattaa 1200
atccatacta ggagcgtcag tgagggtttt catagctttt ggaaatactt tggtctctga 1260
actgtaatta gcaagaagta aaaacagaaa cgtcaaacgt caaatgtttg ctttgttacc 1320
tggaggacta aatgtagatg tctttagtat actttgtatg ttcttaatat tggaagataa 1380
ttttgtgaat ctgtagattt tattttttca gtcttacctt acaaatttct tttctatgaa 1440
taatagagga acttacggca ctctgccatt tgttaatgaa aggaagtgca gaggatttag 1500
aaaagtacat gatccccaga ccacaacaaa ccaaaacata aactcatgtc tgtgtcccat 1560
ggtcatagtc aaagattttg tactgctaaa attaccaaat aatttaaata aagtggattt 1620
gaacacaatt tgaa 1634

<210> 66
<211> 1631
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512931CB1

<400> 66
agcacttgtt ggcctacaga ggtgtggcaa gcagagcacc tcagaactca ggcgtactgc 60
ccgccgcccg agccctgcga gggccgatag cgagggtgtg gcccttatct gcacccagca 120
gagcgccggc ggggtacggt cttaggacct cgatctcctt ctccctcatt ttctctcatc 180
cctacctatt gtgggtttcc gcatgggccg gaccgtggtc gtgctggcgg gaggcatcag 240
cggcttggcc gccagttacc acctgagccg ggccccctgc ccccctaagg tggtcctagt 300
ggagagcagt gagcgtctgg gaggctggat tcgctccgtt cgaggcccta atggtgctat 360
ctttgagctt ggacctcggg gaattaggcc agcgggagcc ctaggggccc ggaccttgct 420
cctggtttct gagcttggct tggattcaga agtgctgcct gtccggggag accacccagc 480
tgcccagaac aggttcctct acgtgggcgg tgccctgcat gccctaccca ctggcctcag 540
ggagctgacc aagccccggg gcaaagagcc tgatgagact gtgcacagtt ttgcccagcg 600
ccgccttgga cctgaggtgg cgtctctagc catggacagt ctctgccgtg gagtgtttgc 660
aggcaacagc cgtgagctca gcatcaggtc ctgctttccc agtctcttcc aagctgagca 720
aacccatcgt tccatattac tgggcctgct gctgggggca gggcggaccc cacagccaga 780
ctcagcactc attcgccagg ccttggctga gcgctggagc cagtggtcac ttcgtggagg 840
tctagagatg ttgcctcagg cccttgaaac ccacctgact agtagggggg tcagtgttct 900
cagaggccag ccggtctgtg ggctcagcct ccaggcagaa gggcgctgga aggtatctct 960
aagggacagc agtctggagg ctgaccacgt tattagtgcc attccagctt cagtgctcag 1020

tgagctgctc cctgctgagg ctgcccctct ggctcgtgcc ctgagtgcca tcactgcagt 1080
gtctgtagct gtggtgaatc tgcagtacca aggagcccat ctgcctgtcc agggatttgg 1140
acatttggtg ccatcttcag aagatccagg agtcctggga atcgtgtatg actcagttgc 1200
tttccctgag caggacggga gcccccctgg cctcagagtg actgtgatgc tgggaggttc 1260
ctggttacag acactggagg ctagtggctg tgtcttatct caggagctgt ttcaacagcg 1320
ggcccaggaa gcagctgcta cacaattagg actgaaggag atgccgagcc actgcttggt 1380
ccatctacac aagaactgca ttccccagta tacactaggt cactggcaaa aactagagtc 1440
agctaggcaa ttcctgactg ctcacaggtt gcccctgact ctggctggag cctcctatga 1500
gggagttgct gttaatgact gtatagagag tgggcgccag gcagcagtca gtgtcctggg 1560
cacagaacct aacagctgat ccccaactct cattcatgaa aataaaaatt gctggagctt 1620
ggcaaaaaaa a 1631

<210> 67
<211> 808
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512933CB1

<400> 67
agcacttgtt ggcctacaga ggtgtggcaa gcagagcacc tcagaactca ggcgtactgc 60
ccgccgcccg agccctgcga gggccgatag cgagggtgtg gcccttatct gcacccagca 120
gagcgccggc ggggtacggt cttaggacct cgatctcctt ctccctcatt ttctctcatc 180
cctacctatt gtgggtttcc gcatgggccg gaccgtggtc gtgctgggcg gaggcatcag 240
cggcttggcc gccagttacc acctgagccg ggccccctgc ccccctaagg tggtcctagt 300
ggagagcagt gagcgtctgg gaggctggat tcgctccgtt cgaggcccta atggtgctat 360
ctttgagctt ggacctcggg gaattaggcc agcgggagcc ctaggggccc ggaccttgct 420
cctggtgatg ctgggaggtt cctggttaca gacactggag gctagtggct gtgtcttatc 480
tcaggagctg tttcaacagc gggcccagga agcagctgct acacaattag gactgaagga 540
gatgccgagc cactgcttgg tccatctaca caagaactgc attccccagt atacactagg 600
tcactggcaa aaactagagt cagctaggca attcctgact gctcacaggt tgcccctgac 660
tctggctgga gcctcctatg agggagttgc tgttaatgac tgtatagaga gtgggcgcca 720
ggcagcagtc agtgtcctgg gcacagaacc taacagctga tccccaactc tcattcatga 780
aaataaaaat tgctggagct tgaaaaaa 808

<210> 68
<211> 657
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512942CB1

<400> 68
cagaggtgtg gcaagcagag cacctcagaa ctcaggcgta ctgcccgccg cccgagccct 60
gcgagggccg atagcgaggg tgtggccctt atctgcaccc agcagagcgc cggcggggta 120
cggtcttagg acctcgatct ccttctccct cattttctct catccctacc tattgtgggt 180
ttccgcatgg gccggaccgt ggtcgtgctg ggcggaggca tcagcggctt ggccgccagt 240
taccacctga gccgggcccc ctgcccccct aaggtgatgc tgggaggttc ctggttacag 300
acactggagg ctagtggctg tgtcttatct caggagctgt ttcaacagcg ggcccaggaa 360
gcagctgcta cacaattagg actgaaggag atgccgagcc actgcttggt ccatctacac 420
aagaactgca ttccccagta tacactaggt cactggcaaa aactagagtc agctaggcaa 480
ttcctgactg ctcacaggtt gcccctgact ctggctggag cctcctatga gggagttgct 540

gttaatgact gtatagagag tgggcgccag gcagcagtca gtgtcctggg cacagaacct 600
aacagctgat ccccaactct cattcatgaa aataaaaatt gctggagctt gaaaaaa 657

<210> 69
<211> 1158
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7513736CB1

<400> 69
ctgaggactg gaagacatta gaataaggtc cagaaatgtc cttggtgtgt ctgacagact 60
ttcaggccca tgcgcgagag cagctgtcta agtcaactcg ggattttatt gaaggtggag 120
cagatgacag catcacgcgg gatgacaaca ttgcagcatt taaaagaatt cgcctccgtc 180
cgcggtacct gagagatgtg tctgaggtgg acaccagaac cacaatccaa ggggaggaga 240
tcagtgcccc tatttgtatc gcacccacag ggttccactg ccttgtctgg cctgatgggg 300
aaatgagcac agcaagagct gcccaagcgg ctggtatctg ctacatcacc agcacatttg 360
ccagctgtag ccttgaagac attgtcattg cagctcccga aggcctccga tggttccaac 420
tctatgtgca tccagacctg cagctgaaca aacagttgat ccagagggta gaatccctag 480
gtttcaaagc tttggtaata actttggata cacctgtatg tggcaacagg cgacatgaca 540
ttcgaaacca gttgaggagg aacttaacac taacagatct tcaatcacct aaaaagattg 600
atgctttgac agaagtggtg gctgctgtaa aggggaaaat tgaagtctac ctggatggcg 660
gggtccgaac tggcaatgat gtgctgaagg ctctggccct tggagctaag tgcatttttc 720
ttgggagacc aatcctatgg ggccttgcct gcaagggtga acatggtgtt aaggaagttt 780
tgaacatttt aacaaatgag ttccacactt ccatggccct tacaggctgc cggtcggtcg 840
ctgagatcaa tcgaaacttg gtccagtttt ccaggctgta agaaaaaagg gccaataacc 900
agactgctga ggttgcccac aggaggatca caaactcaca gcacagtgtg tgatgctgtc 960
cttcctggac cccattctgt ccggaggctc atggcccata tttcccacat ttctaatacc 1020
accacccctg tgcttcaggc cctccaaccc ctgtgttccc caaatgttcc atgcccttct 1080
ttgtatcact gactattata tgttgctctc ttgcctaaat cttcctctga agtaaagatc 1140
tcaaaaggac agatcatt 1158

<210> 70
<211> 2125
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512779CB1

<400> 70
tcttggtcct gctgtcctga ggagctacct gtgcaccctg cctccggctc tcctgagcag 60
agagatcctg atggctgact cagaagcact cccctccctt gctggggacc cagtggctgt 120
ggaagccttg ctccgggccg tgtttggggt tgttgtggat gaggccattc agaaaggaac 180
cagtgtctcc cagaaggtct gtgagtggaa ggagcctgag gagctgaagc agctgctgga 240
tttggagctg cggagccagg gcgagtcaca gaagcagatc ctggagcggt gtcgggctgt 300
gattcgctac agtgtcaaga ctggtcaccc tcggttcttc aaccagctct tctctgggtt 360
ggatccccat gctctggccg ggcgcattat cactgagagc ctcaacacca gccagtacac 420
atatgaaatc gcccccgtgt ttgtgctcat ggaagaggag gtgctgagga aactgcgggc 480
cctggtgggc tggagctctg gggacggaat cttctgccct ggtggctcca tctccaacat 540
gtatgctgta aatctggccc gctatcagcg ctacccggat tgcaagcaga ggggcctccg 600
cacactgccg cccctggccc tattcacatc gaaggagtgt cactactcca tccagaaggg 660
agctgcgttt ctgggacttg gcaccgacag tgtccgagtg gtcaaggctg atgagagagg 720

gaaaatggtc cccgaggatc tggagaggca gattggtatg gccgaggctg agggtgctgt 780
gccgttcctg gtcagtgcca cctctggcac cactgtgcta ggggcctttg accccctgga 840
ggcaattgct gatgtgtgcc agcgtcatgg gctatggctg catgtggatg ctgcctgggg 900
tgggagcgtc ctgctgtcac agacacacag gcatctcctg gatgggatcc agagggctga 960
ctctgtggcc tggaatcccc acaagctcct cgcagcaggc ctgcaatgct ctgcacttct 1020
tctccaggat acctcgaacc tgctcaagcg ctgccatggg tcccaggcca gctacctttt 1080
ccagcaggac aagttctacg atgtggctct ggacacggga gacaaggtgg tgcagtgtgg 1140
ccgccgtgtg gactgtctga agctgtggct catgtggaag gcacagggcg atcaagggct 1200
ggagcggcgc atcgaccagg cctttgtcct tgcccggtac ctggtggagg aaatgaagaa 1260
gcgggaaggg tttgagctag tcatggagcc tgcgagggaa gcaggagagt ccagattacc 1320
acgaaaggct gtcaaaggtg gcccccgtgc tcaaggagcg catggtgaag gagggctcca 1380
tgatgattgg ctaccagccc cacgggaccc ggggcaactt cttccgtgtg gttgtggcca 1440
actctgcact gacctgtgct gatatggact tcctcctcaa cgagctggag cggctaggcc 1500
aggacctgtg agccttctct gtcttgctgc cggccttgat accacccctc acccgcagag 1560
tcactgcatt ccctcccagc ctttgaggcc gggtgcagtg gctcacgcct gtaatcccag 1620
cactttggga ggccgaggcg ggtggatcac ttgaggtcag gagttcgaga ccagcctggc 1680
caataaggtg aaaccctgtc tctactaaaa atacaaaaat tagccgagca tggtggcctg 1740
tgcctgtaaa cccagctact caggaggttg gggcagaatt gcttgaaccc agggggcaga 1800
ggttgcagtg agccgagatt gcacccctgc actccaggct gggcaacagt acgagactct 1860
gttccaaaaa aaataaaaaa gaacagcctt tttaggccac agtgacctgc gcaatgttta 1920
tatgctttga cctactaact ttctcctacc taaatatatg gttttaggag agtgtttaaa 1980
taaattacag tatgtctata tgatggaatg ctcgtttgcc aattgaaatt tgtttccaag 2040
tataaaaaaa aaaaaggggg cccccaatta aaagccacag cccgggatta attcggcccg 2100
taaataagga gggattaaaa tttaa 2125

<210> 71
<211> 1913
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512877CB1

<400> 71
aggttccaag agaaacttag cacttcttgt tctgtttgta gcctcaaatc aacatgtaac 60
caaaagctgt gaatgaataa tccaaacttg aagaacccca gattttattc aggaggagta 120
actggaagcc ataaccatga acttccatat ctgtcatagg aaagaggttg gcctcagcag 180
caatgaggga taattctttc aagaaaggaa tgaagagaga agttagatgg aaaagagatt 240
tttaaggtca aaagaagcag ctgaagaaca tgtgtaggac agccttagga gttactacac 300
ttcaagacct gggtgtctgc tggaggaagg aacatgaccc tgcggagatg ttaagggttg 360
ggagagctga gcttggaggc tggaagccca gctgggaggc tattgttaaa gaggttgctg 420
cagcaactgg tcacacagta gtgttggtag accagacaga ggacatcctg gcaaaatcca 480
aaaagggaat tgaggaaagc cttaggaaag tggcaaagaa gaagtttgca gaaaacccta 540
aggccggcga tgaatttgtg gagaagaccc tgagcaccat agcgaccagc acggatgcag 600
cctccgttgt ccacagcaca gacttggtgg tggaagccat cgtggagaat ctgaaggtga 660
aaaacgagct cttcaaaagg ctggacaagt ttgctgctga gtcattaaaa caccaatgac 720
cagccagaag acatttgaat ctttggtaga ctttagcaaa gccctaggaa agcatcctgt 780
ttcttgcaag gacactcctg ggtttattgt gaaccgcctc ctggttccat acctcatgga 840
agcaatcagg ctgtatgaac gaggtgacgc atccaaagaa gacattgaca ctgctatgaa 900
attaggagcc ggttacccca tgggcccatt tgagcttcta gattatgtcg gactggatac 960
tacgaagttc atcgtggatg ggtggcatga aatggatgca gagaacccat tacatcagcc 1020
cagcccatcc ttaaataagc tggtagcaga gaacaagttc ggcaagaaga ctggagaagg 1080
attttacaaa tacaagtgat gtgcagcttc tccggctctg agaagaacac ctgagagcgc 1140
tttccagcca gtgccccgag tgcctgtggg aatgctcttt ggtcagacat tccctcacac 1200
agtacagttt aataaatgtg cattttgatt gtaatctatc gaagtgatta ttacaccagt 1260

tacagcagta atagattctc cattaagaaa taattccctt ttttagtctg ttcatttctg 1320
tgtattttct aaacagcttt acacccttgg tgccttggag caaacatgtt ttttgaacct 1380
tgtcattttt gtgaagaatt gcctagattc cttctctcat caacgggaaa gtacttcctc 1440
tgagagtgcg agtgcaccat gctcactgtt gctgcgtggg agagtcacaa gccactggca 1500
agcaagtggt atagtctgtg aagcactgca gcgagcagca cctggatctt gcctttataa 1560
gaacatttta ctacctgcag ctttgagtct tgccctacat tttgggcatg acataagatg 1620
tgtctttatt cagctcgtcg tgaagatgct gctgctgaat gggtcagcat atctctgttt 1680
gcatggtttg caggaggtcg gttttcatgg tcattcagtt ccacagatct gaatgattac 1740
tgtctgtctg tgtctttttt ccatgagaaa tcactgttgc aaattgccta taaattgact 1800
ctactaaaat aacaatgttt cagtctgaaa atttgaattg aaaaaaatgt ataatataaa 1860
attgtaatac actcaaatga ttataaaagt aaaagttggt aatttaggca gaa 1913

<210> 72
<211> 1183
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512782CB1

<400> 72
cctgagcagc caacacacca gcccagacag ctgcaagtca ccatggacgc tgaaggcctg 60
gcgctgctgc tgccgcccgt caccctggca gccctggtgg acagctggct ccgagaggac 120
tgcccagggc tcaactacgc agccttggtc agcggggcag gcccctcgca ggcggcgctg 180
tgggccaaat cccctggggt actggcaggg cagcctttct tcgatgccat atttacccaa 240
ctcaactgcc aagtctcctg gttcctcccc gagggatcga agctggtgcc ggtggccaga 300
gtggccgagg cggtgcgggc ggccagacag gcggctgact tcgctctgaa ggtggaagtg 360
gaatgcagca gcctgcagga ggccgtgcag gcagctgagg ctggtgccga ccttgtcctg 420
ctggacaact tcaagccaga ggagctgcac cccacggcca ccgtgctgaa ggcccagttc 480
ccgagtgtgg ctgtggaagc cagtgggggc atcaccctgg acaacctccc ccagttctgc 540
gggccgcaca tagacgtcat ctccatgggg atgctgaccc aggcggcccc agcccttgat 600
ttctccctca agctgtttgc caaagaggtg gctccagtgc ccaaaatcca ctagtcctaa 660
accggaagag gatgacaccg gccatgggtt aacgtggctc ctcaggaccc tctgggtcac 720
acatctttag ggtcagtggc caatggggca catttggcac tagcttgagc ccaactctgg 780
ctctgccacc tgctgctcct gtgacctgtc agggctgact tcacctctgc tcatctcagt 840
ttcctaatct gtaaaatggg tctaataaag gatcaaccac aaaaaaaaaa aaaaaaaaaa 900
aaaaaaaaac aaaaaaacaa aaaaaaaaaa aaacaaaaaa aaaaaaaaaa acaaggggag 960
agcccccccca caggacaccc ctccacaccg agaaataatt agtcagagga gagagcctaa 1020
cgacaagagc agacaccaca gctgcctcca atatgggggg cttgcttaaa aaagcgaggc 1080
gaggcccaaa cagcggacca cacagcaaga tatgcccgag ggaacgaaag gggggcaca 1140
gcgactagga gatcgccaaa acaatagtag aaaacacatg acc 1183

<210> 73
<211> 1074
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512784CB1

<400> 73
gtcctgagca gccaacacac cagcccagac agctgcaagt caccatggac gctgaaggcc 60
tggcgctgct gctgccgccc gtcaccctgg cagccctggt ggacagctgg ctccgagagg 120
actgcccagg gctcaactac gcagccttgg tcagcggggc aggcccctcg caggcggcgc 180

tgtgggccaa atcccctggg gcggtgcggg cggccagaca ggcggctgac ttcgctctga 240
aggtggaagt ggaatgcagc agcctgcagg aggccgtgca ggcagctgag gctggtgccg 300
accttgtcct gctggacaac ttcaagccag aggagctgca ccccacggcc accgtgctga 360
aggcccagtt cccgagtgtg gctgtggaag ccagtggggg catcaccctg gacaacctcc 420
cccagttctg cgggccgcac atagacgtca tctccatggg gatgctgacc caggcggccc 480
cagcccttga tttctccctc aagctgtttg ccaaagaggt ggctccagtg cccaaaatcc 540
actagtccta aaccggaaga ggatgacacc ggccatgggt taacgtggct cctcaggacc 600
ctctgggtca cacatcttta gggtcagtgg ccaatggggc acatttggca ctagcttgag 660
cccaactctg gctctgccac ctgctgctcc tgtgacctgt cagggctgac ttcacctctg 720
ctcatctcag tttcctaatc tgtaaaatgg gtctaataaa ggatcaacca caaaaaaaaa 780
aaaaaaaaaa aaaaaaaaaa caaaaaaaca aaaaaaaaaa aaaacaaaaa aaaaaaaaaa 840
aacaagggga gagccccccc acaggacacc cctccacacc gagaaataat tagtcagagg 900
agagagccta acgacaagag cagacaccac agctgcctcc aatatggggg gcttgcttaa 960
aaaagcgagg cgaggcccaa acagcggacc acacagcaag atatgcccga gggaacgaaa 1020
gggggggcac agcgactagg agatcgccaa aacaatagta gaaaacacat gacc 1074

<210> 74
<211> 1021
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512794CB1

<400> 74
cggcgacatt agctagcgct cgctctactc tctctaacgg gaaagcagcg gaatacaaga 60
gactgaactg tatctgcctc tatttccaaa agactcacgt tcaactttcg ctcacacaaa 120
gccgggaaaa ttttattagt cctttttta aaaaagtta atataaaatt atagcaaaaa 180
aaaaaaggaa cctgaacttt agtaacacag ctggaacaat ccgcagcggc ggcggcagcg 240
gcgggagaag aggtttaatt tagttgattt tctgtggttg ttggttgttc gctagtctca 300
cggtgatgga agctgcacat tttttcgaag ggaccgagaa gctgctggag gtttggttct 360
cccggcagca gcccgacgca aaccaaggat ctggggatct tcgcactatc ccaagatgga 420
gcagcatatt gtatgggacg tatgaattct gactgttggt acttatatac tctggatttc 480
ccagagagtc gggtaatcag tcagccagat caaaccttgg aaattctgat gagtgagctt 540
gacccagcag ttatggacca gttctacatg aaagatggtg ttactgcaaa ggatgtcact 600
cgtgagagtg gaattcgtga cctgatacca ggttctgtca ttgatgccac aatgttcaat 660
ccttgtgggt attcgatgaa tggaatgaaa tcggatggaa cttattggac tattcacatc 720
actccagaac cagaattttc ttatgttagc tttgaaacaa acttaagtca gacctcctat 780
gatgacctga tcaggaaagt tgtagaagtc ttcaagccag gaaaatttgt gaccaccttg 840
tttgttaatc agagttctaa atgtcgcaca gtgcttgctt cgccccagaa gattgaaggt 900
tttaagcgtc ttgattgcca gagtgctatg ttcaatgatt acaattttgt ttttaccagt 960
tttgctaaga agcagcaaca acagcagagt tgattaagaa aaatgaagaa aaaacgcaaa 1020
a 1021

<210> 75
<211> 1127
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512886CB1

<400> 75
agagagagag agagagaact agtctcgagc catcatggaa gcaatgtggc tcctgtgtgt 60

ggcgttggcg gtcttggcat ggggcttcct ctgggtttgg gactcctcag aacgaatgaa 120
gagtcgggag cagggaggac ggctgggagc cgaaagccgg accctgctgg tcatagcgca 180
ccctgacgat gaagccatgt tttttgctcc cacagtgcta ggcttggccc gcctaaggca 240
ctgggtgtac ctgctttgct tctctgcagg aaattactac aatcaaggag agactcgtaa 300
gaaagaactt ttgcagagct gtgatgtttt ggggattcca ctctccagtg taatgattat 360
tgacaacagg gatttcccag atgacccagg catgcagtgg gacacagagc acgtggccag 420
agtcctcctt cagcacatag aagtgaatgg catcaatctg gtggtgactt tcgatgcagg 480
gggagtaagt ggccacagca atcacattgc tctgtatgca gctgtgaggg tgctctgtgc 540
tcacgcttca gtctgtgaat gtgctgcgca agtacatctc ccttctggat ctgcccttgt 600
ctctgcttca tacgcaggat gtcctcttcg tgctcaacag caaagaagtg gcacaggcca 660
agaaagccat gtcctgccac cgcagccagc tcctctggtt ccgccgcctc tacattatct 720
tctcccggta catgagaatc aactcactga gcttcctctg aagccttgaa gggttttcag 780
atccaaggaa caaaggggaa aatagacaaa ggagtgcaga ggacctggcc tggcactggc 840
ttatttacct gagctcaagg agatccccgc tggagcagcc tctgcaaaag ggagcccatg 900
taggccaggg gctgtccaaa ctccagcttc ttcccctggg aaaaaaccca aagaaccaaa 960
aacaaaccac cccaaggata ataatagcta cactgctagc ttctcaagtt cttgtgaaaa 1020
acaatttaca taatgacaca gtagatgtgg aacacctagc ccagtgcctg ggcaggtccc 1080
tattatcata aatgaacata aaagtgctct aaaaacactc aaaaaaa 1127

<210> 76
<211> 1855
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512929CB1

<400> 76
caggcgggcc cccgcgcggc agggccctgg acccgcgcgg ctcccgggga tggtgagcaa 60
ggcgctgctg cgcctcgtgt ctgccgtcaa ccgcaggagg atgaagctgc tgctgggcat 120
cgccttgctg gcctacgtcg cctctgtttg gggcaacttc gttaatatga ggtctatcca 180
ggaaaatggt gaactaaaaa ttgaaagcaa gattgaagag atggttgaac cactaagaga 240
gaaaatcaga gatttagaaa aaagctttac ccagaaatac ccaccagtaa agtttttatc 300
agaaaaggat cggaaaagaa ttttgttgac cagatatacc atctggcatc tccagcctcc 360
cctccaaact acatgtataa tcctatcaag acattaaaga ccaatacgat tgggacatta 420
aacatgttgg ggctggcaaa acgagtcggt gcccgtctgc tcctggcctc cacatcggag 480
gtgtatggag atcctgaagt ccaccctcaa agtgaggatt actggggcca cgtgaatcca 540
ataggacctc gggcctgcta cgatgaaggc aaacgtgttg cagagaccat gtgctatgcc 600
tacatgaagc aggaaggcgt ggaagtgcga gtggccagaa tcttcaacac ctttgggcca 660
cgcatgcaca tgaacgatgg gcgagtagtc agcaacttca tcctgcaggc gctccagggg 720
gagccactca cggtatacgg atccgggtct cagacaaggg cgttccagta cgtcagcgat 780
ctagtgaatg gcctcgtggc tctcatgaac agcaacgtca gcagcccggt caacctgggg 840
aacccagaag aacacacaat cctagaattt gctcagttaa ttaaaaacct tgttggtagc 900
ggaagtgaaa ttcagtttct ctccgaagcc caggatgacc cacagaaaag aaaaccagac 960
atcaaaaaag caaagctgat gctggggtgg gagcccgtgg tcccgctgga ggaaggttta 1020
aacaaagcaa ttcactactt ccgtaaagaa ctcgagtacc aggcaaataa tcagtacatc 1080
cccaaaccaa agcctgccag aataaagaaa ggacggactc gccacagctg aactcctcac 1140
ttttaggaca caagactacc attgtacact tgatgggatg tatttttggc tttttttgt 1200
tgtcgtttaa agaaagactt taacaggtgt catgaagaac aaactggaat ttcattctga 1260
agcttgcttt aatgaaatgg atgtgcctaa aagctcccct caaaaaactg cagattttgc 1320
cttgcacttt ttgaatctct ctttttatgt aaaatagcgt agatgcatct ctgcgtattt 1380
tcaagttttt ttatcttgct gtgagagcat atgttgtgac tgtcgttgac agttttattt 1440
actggtttct ttgtgaagct gaaaaggaac attaagcggg acaaaaaatg ccgattttat 1500
ttataaaagt gggtacttaa taaatgagtc gttatactat gcataaagaa aaatcctagc 1560
agtattgtca ggtggtggtg cgccggcatt gattttaggg cagataaaag aattctgtgt 1620

```
gagagcttta tgtttctctt ttaattcaga gtttttccaa ggtctacttt tgagttgcaa 1680
acttgacttt gaaatattcc tgttggtcat gatcaaggat atttgaaatc actactgtgt 1740
tttgctgcgt atctggggcg ggggcaggtt gggggcaca aagttaacat attcttggtt 1800
aaccatggtt aaatatgcta ttttaataaa atattgaaac tcaccaaaaa aaaaa      1855
```